

# Confidence Bands for Distribution Functions: A New Look at the Law of the Iterated Logarithm

Lutz Dümbgen\* and Jon A. Wellner\*\*

March 2014

## Abstract

We present a general law of the iterated logarithm for stochastic processes on the open unit interval having subexponential tails in a locally uniform fashion. It applies to standard Brownian bridge but also to suitably standardized empirical distribution functions. This leads to new goodness-of-fit tests and confidence bands which refine the procedures of Berk and Jones (1979) and Owen (1995). Roughly speaking, the high power and accuracy of the latter procedures in the tail regions of distributions are essentially preserved while gaining considerably in the central region.

\*Work supported by Swiss National Science Foundation.

\*\*Work supported in part by NSF DMS 11-04832 and NI-AID Grant 2R01 AI291968-04.

**AMS subject classifications.** 62E20, 62G15, 62G30.

**Key words.** Confidence band, limit distribution, sub-exponential tails, submartingale, tail regions.

## 1 Introduction

Let  $\hat{F}_n$  be the empirical distribution function of independent random variables  $X_1, X_2, \dots, X_n$  with unknown distribution function  $F$  on the real line. Let us recall some well-known facts about  $\hat{F}_n$  (cf. Shorack and Wellner 1986): The stochastic process  $(\hat{F}_n(x))_{x \in \mathbb{R}}$  has the same distribution as  $(\hat{G}_n(F(x)))_{x \in \mathbb{R}}$ , where  $\hat{G}_n$  is the empirical distribution of independent random variables  $U_1, U_2, \dots, U_n$  with uniform distribution on  $[0, 1]$ . This enables us to construct confidence bands for the distribution function  $F$ . A well-known classical method are Kolmogorov-Smirnov confidence bands: Let

$$\mathbb{U}_n(t) := n^{1/2}(\hat{G}_n(t) - t),$$

and let  $\kappa_{n,\alpha}^{\text{KS}}$  be the  $(1 - \alpha)$ -quantile of

$$\|\mathbb{U}_n\|_\infty := \sup_{t \in [0,1]} |\mathbb{U}_n(t)|.$$

Then with probability at least  $1 - \alpha$ ,

$$F(x) \in [\widehat{F}_n(x) \pm n^{-1/2} \kappa_{n,\alpha}^{\text{KS}}] \quad \text{for all } x \in \mathbb{R}. \quad (1)$$

Equality holds if  $F$  is continuous. Since  $\mathbb{U}_n$  converges in distribution in  $\ell_\infty([0, 1])$  to standard Brownian bridge  $\mathbb{U}$ ,  $\kappa_{n,\alpha}^{\text{KS}}$  converges to the  $(1 - \alpha)$ -quantile  $\kappa_\alpha^{\text{KS}}$  of  $\|\mathbb{U}\|_\infty$ . In particular, the simultaneous confidence intervals in (1) have width  $O(n^{-1/2})$  uniformly in  $x \in \mathbb{R}$ .

Another method, based on a goodness-of-fit test by Berk and Jones (1979), was introduced by Owen (1995): Let  $\kappa_{n,\alpha}^{\text{BJ}}$  be the  $(1 - \alpha)$ -quantile of

$$T_n^{\text{BJ}} := n \sup_{t \in (0,1)} K(\widehat{G}_n(t), t),$$

where

$$K(s, t) := s \log \frac{s}{t} + (1 - s) \log \frac{1 - s}{1 - t}$$

for  $s \in [0, 1]$  and  $t \in (0, 1)$ . This leads to an alternative confidence band for  $F$ : With probability at least  $1 - \alpha$ ,

$$nK(\widehat{F}_n(x), F(x)) \leq \kappa_{n,\alpha}^{\text{BJ}} \quad \text{for all } x \in \mathbb{R}. \quad (2)$$

As shown by Jager and Wellner (2007), the asymptotic distribution of  $T_n^{\text{BJ}}$  remains the same if one replaces  $K(s, t)$  by a more general function; in particular, one may interchange its two arguments. Moreover,

$$\kappa_{n,\alpha}^{\text{BJ}} = \log \log(n) + 2^{-1} \log \log \log(n) + O(1).$$

From this one can deduce that (2) leads to confidence intervals with length at most

$$2(2\gamma_n F(x)(1 - F(x)))^{1/2} + 2\gamma_n \quad \text{where} \quad \gamma_n := \frac{\kappa_{n,\alpha}^{\text{BJ}}}{n} = (1 + o(1)) \frac{\log \log n}{n},$$

uniformly in  $x \in \mathbb{R}$ ; see (K.5) in Section 6.2. Hence they are substantially shorter than the Kolmogorov-Smirnov intervals for  $F(x)$  close to 0 or 1. But in the central region, i.e. when  $F(x)$  is bounded away from 0 and 1, they are of width  $O(n^{-1/2}(\log \log n)^{1/2})$  rather than  $O(n^{-1/2})$ . An obvious goal is to refine these methods and combine the benefits of the Kolmogorov-Smirnov and Berk-Jones confidence bands. Methods of this type have been proposed by various authors, see Mason and Schuenemeyer (1983) and the references cited therein.

A key for understanding the asymptotics of  $T_n^{\text{BJ}}$  but also the new methods presented later are suitable variants of the law of the iterated logarithm (LIL). For Brownian bridge  $\mathbb{U}$  the LIL states that

$$\limsup_{t \downarrow 0} \frac{\mathbb{U}(t)}{\sqrt{2t \log \log(1/t)}} = \limsup_{t \uparrow 1} \frac{\mathbb{U}(t)}{\sqrt{2(1-t) \log \log(1/(1-t))}} = 1 \quad (3)$$

almost surely. Various refinements of this result have been obtained. One particular consequence of Kolmogorov's upper class test (cf. Erdős 1942, or Ito and McKean 1974, Chapter 1.8) is the following result: For  $t \in (0, 1)$  define

$$\begin{aligned} C(t) &:= \log \log \frac{e}{4t(1-t)} = \log(1 - \log(1 - (2t-1)^2)) \geq 0, \\ D(t) &:= \log(1 + C(t)^2) \geq 0. \end{aligned}$$

Then for any fixed  $\nu > 3/4$ ,

$$\sup_{t \in (0,1)} \left( \frac{\mathbb{U}(t)^2}{2t(1-t)} - C(t) - \nu D(t) \right) < \infty \quad (4)$$

almost surely. Note that  $C(t) = C(1-t)$ ,  $D(t) = D(1-t)$ , and, as  $t \downarrow 0$ ,

$$\begin{aligned} C(t) &= \log \log(1/t) + O(\log(1/t)^{-1}), \\ D(t) &= 2 \log \log \log(1/t) + O((\log \log(1/t))^{-1}). \end{aligned}$$

This explains why (4) follows from Kolmogorov's test and shows the connection between (4) and (3). Note also that

$$\lim_{t \rightarrow 1/2} \frac{C(t)}{(2t-1)^2} = \lim_{t \rightarrow 1/2} \frac{D(t)}{(2t-1)^4} = 1.$$

In the present paper we prove statements similar to (4) for general stochastic processes on  $(0, 1)$ . In Section 2 we state a general condition on a stochastic process  $X = (X(t))_{t \in (0,1)}$  such that for any fixed  $\nu > 1$ ,

$$\sup_{t \in (0,1)} (X(t) - C(t) - \nu D(t)) < \infty$$

almost surely. In particular, the stochastic process

$$X(t) := \frac{\mathbb{U}(t)^2}{2t(1-t)}$$

satisfies this condition. Then in Section 3 these general results are applied to

$$X_n(t) := nK(\hat{G}_n(t), t).$$

It turns out that for any fixed  $\nu > 1$ ,

$$T_{n,\nu} := \sup_{t \in (0,1)} (nK(\hat{G}_n(t), t) - C(t) - \nu D(t)) \quad (5)$$

converges in distribution to

$$T_\nu := \sup_{t \in (0,1)} \left( \frac{\mathbb{U}(t)^2}{2t(1-t)} - C(t) - \nu D(t) \right).$$

Asymptotic statements like this refer to  $n \rightarrow \infty$ , unless stated otherwise. Moreover, if  $U_{n:1} < U_{n:2} < \dots < U_{n:n}$  are the order statistics of  $U_1, U_2, \dots, U_n$ , then for fixed  $\nu > 1$ ,

$$\tilde{T}_{n,\nu} := \max_{j=1,2,\dots,n} ((n+1)K(t_{nj}, U_{n:j}) - C(t_{nj}) - \nu D(t_{nj})) \rightarrow_{\mathcal{L}} T_\nu,$$

where

$$t_{nj} := \frac{j}{n+1} \quad \text{for } j = 1, 2, \dots, n.$$

To test the null hypothesis that  $F$  is equal to a given continuous distribution function  $F_o$ , consider the test statistic

$$T_{n,\nu}(F_o) := \sup_{x \in \mathbb{R}} (nK(\hat{F}_n(x), F_o(x)) - C(F_o(x)) - \nu D(F_o(x))). \quad (6)$$

Under the null hypothesis,  $T_{n,\nu}(F_o)$  has the same distribution as  $T_{n,\nu}$ . Hence if  $\kappa_{n,\nu,\alpha}$  denotes the  $(1-\alpha)$ -quantile of  $T_{n,\nu}$ , one may reject the null hypothesis at level  $\alpha \in (0, 1)$  if  $T_{n,\nu}(F_o)$  exceeds  $\kappa_{n,\nu,\alpha}$ . In Section 4 we investigate the power of this new test in more detail. In particular we show that it attains the detection boundary for Gaussian mixture models as specified by Donoho and Jin (2004).

The statistic  $\tilde{T}_{n,\nu}$  leads to a new confidence band for  $F$ : Let  $-\infty = X_{n:0} < X_{n:1} \leq X_{n:2} \leq \dots \leq X_{n:n} < X_{n:n+1} = \infty$  be the order statistics of  $X_1, X_2, \dots, X_n$ , and let  $\tilde{\kappa}_{n,\nu,\alpha}$  and  $\kappa_{\nu,\alpha}$  be the  $(1-\alpha)$ -quantile of  $\tilde{T}_{n,\nu}$  and  $T_\nu$ , respectively. Then  $\tilde{\kappa}_{n,\nu,\alpha} \rightarrow \kappa_{\nu,\alpha}$ , and with probability at least  $1-\alpha$ , the following is true: For  $0 \leq j \leq n$  and  $X_{n:j} \leq x < X_{n:j+1}$ ,

$$F(x) \in [a_{nj}, b_{nj}],$$

where  $a_{n0} := 0$ ,  $b_{nn} := 1$  and

$$\begin{aligned} a_{nj} &:= \min\{u \in [0, 1] : nK(t_{nj}, u) \leq C(t_{nj}) + \nu D(t_{nj}) + \tilde{\kappa}_{n,\nu,\alpha}\} \quad \text{if } j > 0, \\ b_{nj} &:= \max\{u \in [0, 1] : nK(t_{n,j+1}, u) \leq C(t_{n,j+1}) + \nu D(t_{n,j+1}) + \tilde{\kappa}_{n,\nu,\alpha}\} \quad \text{if } j < n. \end{aligned}$$

Since  $C(t_{nj}) + \nu D(t_{nj}) + \tilde{\kappa}_{n,\nu,\alpha}$  is no larger than

$$C(t_{n1}) + \nu D(t_{n1}) + \tilde{\kappa}_{n,\nu,\alpha} = (1 + o(1)) \log \log n$$

for  $1 \leq j \leq n$ , our confidence bands have similar accuracy as those of Owen (1995) in the tail regions while achieving the usual root- $n$  consistency everywhere. A more precise comparison is provided in Section 5. Thereafter we relate our methods to a negative result of Bahadur and Savage (1956) about the nonexistence of confidence bands with vanishing width in the tails. Finally we

discuss briefly an interesting alternative approach to goodness-of-fit tests and confidence bands by Aldor-Noiman et al. (2013) and Eiger et al. (2013).

All proofs and technical arguments are deferred to Section 6. Section 7 contains supplementary material including a quantitative version of Bahadur and Savage (1956, Theorem 2) and decision theoretic considerations about the Gaussian mixture model of Donoho and Jin (2004).

## 2 A general non-Gaussian LIL

Our conditions and results involve the function  $\text{logit} : (0, 1) \rightarrow \mathbb{R}$  with

$$\text{logit}(t) := \log\left(\frac{t}{1-t}\right).$$

Its inverse is the logistic function  $\ell : \mathbb{R} \rightarrow (0, 1)$  with

$$\ell(x) := \frac{e^x}{1 + e^x} = \frac{1}{e^{-x} + 1},$$

and

$$\ell'(x) = \ell(x)(1 - \ell(x)) = \frac{1}{e^x + e^{-x} + 2}.$$

We consider stochastic processes  $X = (X(t))_{t \in \mathcal{T}}$  on subsets  $\mathcal{T}$  of  $(0, 1)$  which have locally uniformly sub-exponential tails in the following sense:

**Condition 2.1.** There exist a real constant  $M \geq 1$  and a non-increasing function  $L : [0, \infty) \rightarrow [0, 1]$  such that  $L(c) = 1 - O(c)$  as  $c \downarrow 0$ , and

$$\mathbb{P}\left(\sup_{t \in [\ell(a), \ell(a+c)] \cap \mathcal{T}} X(t) > \eta\right) \leq M \exp(-L(c)\eta) \quad (7)$$

for arbitrary  $a \in \mathbb{R}$ ,  $c \geq 0$  and  $\eta \in \mathbb{R}$ .

**Theorem 2.2.** Suppose that  $X$  satisfies Condition 2.1. For arbitrary  $\nu > 1$  and  $L_o \in (0, 1)$  there exists a real constant  $M_o \geq 1$  depending only on  $M$ ,  $L(\cdot)$ ,  $\nu$  and  $L_o$  such that

$$\mathbb{P}\left(\sup_{t \in \mathcal{T}} (X(t) - C(t) - \nu D(t)) > \eta\right) \leq M_o \exp(-L_o \eta) \quad \text{for arbitrary } \eta \geq 0.$$

**Remark 2.3.** Suppose that  $X$  satisfies Condition 2.1, where  $\inf(\mathcal{T}) = 0$  and  $\sup(\mathcal{T}) = 1$ . For any  $\nu > 1$ , the supremum  $T_\nu(X)$  of  $X - C - \nu D$  over  $\mathcal{T}$  is finite almost surely. But this implies that

$$\lim_{t \rightarrow \{0,1\}} (X(t) - C(t) - \nu D(t)) = -\infty$$

almost surely. For if  $1 < \nu' < \nu$ , then

$$X(t) - C(t) - \nu D(t) \leq T_{\nu'}(X) - (\nu - \nu')D(t),$$

so the claim follows from  $T_{\nu'}(X) < \infty$  almost surely and  $D(t) \rightarrow \infty$  as  $t \rightarrow \{0, 1\}$ .

**Remark 2.4.** Our definition of the function  $D = \log(1 + C^2)$  may look somewhat arbitrary. Indeed, we tried various choices, e.g.  $D = 2 \log(1 + C)$ . Theorem 2.2 is valid for any nonnegative function  $D$  on  $(0, 1)$  such that  $D(1 - \cdot) = D(\cdot)$  and  $D(t)/\log \log \log(1/t) \rightarrow 2$  as  $t \downarrow 0$ . The special choice  $D = \log(1 + C^2)$  yielded a rather uniform distribution of  $\arg \max_{(0,1)} (X - C - \nu D)$  when  $X(t) = \mathbb{U}(t)^2/(2t(1-t))$  and  $\nu$  close to one.

Our first example for a process  $X$  satisfying Condition 2.1 is squared and standardized Brownian bridge:

**Lemma 2.5.** *Let  $\mathcal{T} = (0, 1)$  and  $X(t) = \mathbb{U}(t)^2/(2t(1-t))$  with standard Brownian bridge  $\mathbb{U}$ . Then Condition 2.1 is satisfied with  $M = 2$  and  $L(c) = e^{-c}$ .*

In particular, Lemma 2.5 and Theorem 2.2 yield (4) for any  $\nu > 1$ .

### 3 Implications for the uniform empirical process

As indicated in the introduction, Theorem 2.2 may be applied to the uniform empirical process  $\hat{G}_n$  in two ways. A first version concerns  $\mathcal{T} = (0, 1)$  and

$$X_n(t) := nK(\hat{G}_n(t), t).$$

**Lemma 3.1.** *The stochastic process  $X_n$  satisfies Condition 2.1 with  $M = 2$  and  $L(c) = e^{-c}$ .*

Combining this lemma, Theorem 2.2 and Donsker's Theorem for the uniform empirical process yields the following result:

**Theorem 3.2.** *For any fixed  $\nu > 1$ ,*

$$T_{n,\nu} = \sup_{t \in (0,1)} (X_n(t) - C(t) - \nu D(t))$$

*converges in distribution to the random variable*

$$T_\nu := \sup_{t \in (0,1)} \left( \frac{\mathbb{U}(t)^2}{2t(1-t)} - C(t) - \nu D(t) \right).$$

For the computation of confidence bands it is more convenient to work with the following stochastic process on  $\mathcal{T}_n := \{t_{nj} : j = 1, 2, \dots, n\}$ :

$$\tilde{X}_n(t_{nj}) := (n+1)K(t_{nj}, U_{n:j}).$$

**Lemma 3.3.** *The stochastic process  $\tilde{X}_n$  satisfies Condition 2.1 with  $M = 2$  and  $L(c) = e^{-c}$ .*

Again we may combine this with Theorem 2.2 and Donsker's theorem for partial sum processes to obtain a new limit theorem:

**Theorem 3.4.** *For any fixed  $\nu > 1$ ,*

$$\tilde{T}_{n,\nu} = \sup_{t \in \mathcal{T}_n} (\tilde{X}_n(t) - C(t) - \nu D(t))$$

*converges in distribution to the random variable  $T_\nu$  defined in Theorem 3.2.*

## 4 Goodness-of-fit tests

As explained in the introduction, we may reject the null hypothesis that  $F$  is a given continuous distribution function  $F_o$  at level  $\alpha$  if

$$T_{n,\nu}(F_o) = \sup_{x \in \mathbb{R}} (nK(\hat{F}_n(x), F_o(x)) - C(F_o(x)) - \nu D(F_o(x))),$$

exceeds  $\kappa_{n,\nu,\alpha}$ . Note also that the latter supremum may be expressed as the maximum of  $2n+1$  terms, replacing the argument  $(x)$  with  $(X_{n:i})$  and  $(X_{n:i} -)$  for  $1 \leq i \leq n$  or with  $(F_o^{-1}(1/2))$ .

As shown in the next lemma, for any fixed critical value  $\kappa > 0$ , the probability that  $T_{n,\nu}(F_o) \leq \kappa$  is small if the quantity

$$\Delta_n(F, F_o) := \sup_{\mathbb{R}} \frac{\sqrt{n}|F - F_o|}{\sqrt{\Gamma(F_o)F_o(1 - F_o) + \Gamma(F_o)/\sqrt{n}}}$$

is large, where  $\Gamma(t) := C(t) + 1$  for  $t \in [0, 1]$  with  $C(0) := C(1) := \infty$ . Note that  $\Gamma(t)t(1-t) \rightarrow 0$  as  $(0, 1) \ni t \rightarrow \{0, 1\}$ .

**Lemma 4.1.** *For any critical value  $\kappa > 0$  there exists a constant  $B_{\nu,\kappa}$  such that*

$$\mathbb{P}_F(T_{\nu,n}(F_o) \leq \kappa) \leq B_{\nu,\kappa} \Delta_n(F, F_o)^{-4/5}. \quad (8)$$

Here and subsequently, the subscript  $F$  in  $\mathbb{P}_F(\cdot)$  or  $\mathbb{E}_F(\cdot)$  specifies the true distribution function of the random variables  $X_1, X_2, \dots, X_n$ . Now consider an arbitrary sequence  $(F_n)_n$

of distribution functions. Then for any fixed level  $\alpha \in (0, 1)$ , Lemma 4.1 and the fact that  $\kappa_{n,\nu,\alpha} \rightarrow \kappa_{\nu,\alpha} < \infty$  imply that

$$\mathbb{P}_{F_n}(T_{n,\nu}(F_o) > \kappa_{n,\nu,\alpha}) \rightarrow 1$$

provided that

$$\Delta_n(F_n, F_o) \rightarrow \infty. \quad (9)$$

In particular, (9) is satisfied if  $F_n \equiv F_* \neq F_o$  for all sample sizes  $n$ . Thus our test has asymptotic power one for any fixed distribution function different from  $F_o$ .

**Detecting Gaussian mixtures.** We consider a testing problem studied in detail by Donoho and Jin (2004). The null hypothesis is given by  $F_o = \Phi$ , the standard Gaussian distribution function, whereas

$$F_n(x) := (1 - \varepsilon_n)\Phi(x) + \varepsilon_n\Phi(x - \mu_n)$$

for certain numbers  $\varepsilon_n \in (0, 1)$  and  $\mu_n > 0$ . By means of Lemma 4.1 one can derive the following result:

**Lemma 4.2.** (a) Suppose that  $\varepsilon_n = n^{-\beta+o(1)}$  for some fixed  $\beta \in (1/2, 1)$ . Further let  $\mu_n = \sqrt{2r \log(n)}$  for some  $r \in (0, 1)$ . Then  $\Delta_n(F_n, \Phi) \rightarrow \infty$  if

$$r > r_*(\beta) := \begin{cases} \beta - 1/2 & \text{if } \beta \in (1/2, 3/4], \\ (1 - \sqrt{1 - \beta})^2 & \text{if } \beta \in [3/4, 1). \end{cases}$$

(b) Suppose that  $\varepsilon_n = n^{-1/2+o(1)}$  such that  $\pi_n := \sqrt{n}\varepsilon_n \rightarrow 0$ . Then  $\Delta_n(F_n, \Phi) \rightarrow \infty$  if  $\mu_n = \sqrt{2s \log(1/\pi_n)}$  for some  $s > 1$ .

As explained by Donoho and Jin (2004), any goodness-of-fit test at fixed level  $\alpha \in (0, 1)$  has trivial asymptotic power  $\alpha$  whenever  $\varepsilon_n = n^{-\beta}$  for some  $\beta \in (1/2, 1)$  and  $\mu_n = \sqrt{2r \log(n)}$  with  $r < r_*(\beta)$ . Thus our new test provides another example of an asymptotically optimal procedure in this particular setting. Other procedures with asymptotic power one whenever  $r > r_*(\beta)$  are Tukey's higher criticism test (Donoho and Jin 2004) or the generalized Berk–Jones tests (Jager and Wellner 2007).

In the setting of part (b), the latter two classes of tests can fail to have asymptotic power one if  $\mu_n = \sqrt{2s \log(1/\pi_n)}$  for fixed  $s > 1$  but  $\pi_n \rightarrow 0$  sufficiently slow. On the other hand, one can show that any level- $\alpha$  test of  $F_o$  versus  $F_n$  has trivial asymptotic power whenever  $\mu_n \leq$



$\sqrt{2s \log(1/\pi_n)}$  for an arbitrary fixed  $s < 1$ . A rigorous proof is provided with the supplementary material.

Parts (a) and (b) of Lemma 4.2 are well connected. For let  $\varepsilon_n = n^{-\beta+o(1)}$  for some  $\beta \in (1/2, 3/4]$ , and  $\mu_n = \sqrt{2r \log(n)}$  for some  $r > \beta - 1/2$ . Then  $s := r/(\beta - 1/2) > 1$  and with  $\pi_n := \sqrt{n}\varepsilon_n = n^{1/2-\beta+o(1)}$  we may rewrite  $\mu_n$  as

$$\mu_n = \sqrt{2s(\beta - 1/2) \log(n)} = \sqrt{(2s + o(1)) \log(1/\pi_n)}.$$

## 5 Confidence bands

The confidence bands of Owen (1995) may be described as follows: For  $0 \leq j \leq n$  let  $s_{nj} := j/n$ . With confidence  $1 - \alpha$  we may claim that for  $0 \leq j \leq n$  and  $X_{n:j} \leq x < X_{n:j+1}$ ,

$$F(x) \in [a_{nj}^{\text{BJO}}, b_{nj}^{\text{BJO}}],$$

where

$$\begin{aligned} b_{nj}^{\text{BJO}} &:= \begin{cases} \max\{b \in (s_{nj}, 1) : K(s_{nj}, b) \leq \gamma_n^{\text{BJ}}\} & \text{for } 0 \leq j < n, \\ 1 & \text{for } j = n, \end{cases} \\ a_{nj}^{\text{BJO}} &:= 1 - b_{n, n-j}^{\text{BJO}}, \end{aligned}$$

and

$$\gamma_n^{\text{BJ}} = \frac{\kappa_{n,\alpha}^{\text{BJ}}}{n} = \frac{\log \log n}{n} (1 + o(1)).$$

Our new method is analogous: With confidence  $1 - \alpha$ , for  $0 \leq j \leq n$  and  $X_{n:j} \leq x < X_{n:j+1}$ , the value  $F(x)$  is contained in  $[a_{nj}, b_{nj}]$ , where

$$\begin{aligned} b_{nj} &:= \begin{cases} \max\{u \in (t_{n,j+1}, 1) : K(t_{n,j+1}, u) \leq \gamma_n(t_{n,j+1})\} & \text{for } 0 \leq j < n, \\ 1 & \text{for } j = n, \end{cases} \\ a_{nj} &:= 1 - b_{n, n-j}, \end{aligned}$$

and

$$\gamma_n(t) := \frac{C(t) + \nu D(t) + \tilde{\kappa}_{n,\nu,\alpha}}{n+1}$$

for  $t \in \mathcal{T}_n$ . Asymptotically the new confidence band is everywhere at least as good as Owen's (1995) band, and in the central region it is infinitely more accurate:

**Theorem 5.1.** *For any fixed  $\alpha \in (0, 1)$ ,*

$$\max_{j=0,1,\dots,n-1} \frac{b_{nj} - s_{nj}}{b_{nj}^{\text{BJO}} - s_{nj}} = \max_{j=1,2,\dots,n} \frac{s_{nj} - a_{nj}}{s_{nj} - a_{nj}^{\text{BJO}}} \rightarrow 1,$$

while

$$\begin{aligned}\max_{j=0,1,\dots,n} (b_{nj}^{\text{BJO}} - s_{nj}) &= \max_{j=0,1,\dots,n} (s_{nj} - a_{nj}^{\text{BJO}}) = (1 + o(1)) \sqrt{\frac{\log \log n}{2n}}, \\ \max_{j=0,1,\dots,n} (b_{nj} - s_{nj}) &= \max_{j=0,1,\dots,n} (s_{nj} - a_{nj}) = O(n^{-1/2}).\end{aligned}$$

To be honest, the asymptotic statement in the first part of Theorem 5.1 requires huge sample sizes to materialize. In our numerical experiments it turned out that for sample sizes  $n$  up to 10000 and very small indices  $j$ , the ratio  $(b_{nj} - s_{nj})/(b_{nj}^{\text{BJO}} - s_{nj})$  is between 1.5 and 2 but drops off quickly as  $j$  gets larger.

**Numerical example.** The left panel in Figure 1 depicts for  $n = 500$ ,  $\nu = 1.1$  and  $\alpha = 5\%$  the confidence limits  $a_{nj}$  and  $b_{nj}$  as functions of  $j \in \{0, 1, \dots, n\}$ . The dotted (yellow) line in the middle represents the values  $s_{nj}$ . The corresponding quantile  $\tilde{\kappa}_{n,\nu,\alpha}$  was estimated in 40000 Monte-Carlo simulations as 4.2471, and this leads to the maximal value  $\gamma_n(t_{n1}) = 0.0151$ . In the right panel one sees the centered boundaries  $a_{nj} - s_{nj}$  and  $b_{nj} - s_{nj}$ . In addition the centered boundaries  $a_{nj}^{\text{BJO}} - s_{nj}$  and  $a_{nn}^{\text{BJO}} - s_{nn}$  are shown as dashed (and cyan) lines, based on the estimated quantile  $\kappa_{n,\alpha}^{\text{BJ}} = 5.6615$  and  $\gamma_n^{\text{BJO}} = 0.0113$ . The additional horizontal lines are the values  $\pm n^{-1/2} \kappa_{n,\alpha}^{\text{KS}} = \pm 0.0604$  for the Kolmogorov-Smirnov bands.

Figure 2 shows the same as the right panel in Figure 1, but with sample sizes  $n = 2000$  and  $n = 8000$  in the left and right panel, respectively.

**Accuracy in the tails.** The confidence bands described here yield an upper bound for  $F$  with limit  $b_{n0}^{\text{BJO}}$  or  $b_{n0}$  at  $-\infty$  and a lower bound for  $F$  with limit  $a_{nn}^{\text{BJO}}$  or  $a_{nn}$  at  $+\infty$ . The proof of Theorem 5.1 reveals that

$$b_{n0}^{\text{BJO}}, b_{nn} = \frac{\log \log n}{n} (1 + o(1)) \quad \text{and} \quad a_{nn}^{\text{BJO}}, a_{nn} = 1 - \frac{\log \log n}{n} (1 + o(1)).$$

On the other hand, the proof of Theorem 2 of Bahadur and Savage (1956) shows that we cannot expect substantially more accuracy in the tails. Their arguments can be adapted to show that for any  $(1 - \alpha)$ -confidence band and any  $c > 0$ , the limit of the upper band at  $-\infty$  is smaller than  $c/n$  with probability at most  $(1 - c/n)^{-n} \alpha$ . The same bound holds true for the probability that the limit of the lower band at  $\infty$  is greater than  $1 - c/n$ . For a proof we refer to the supplementary material.

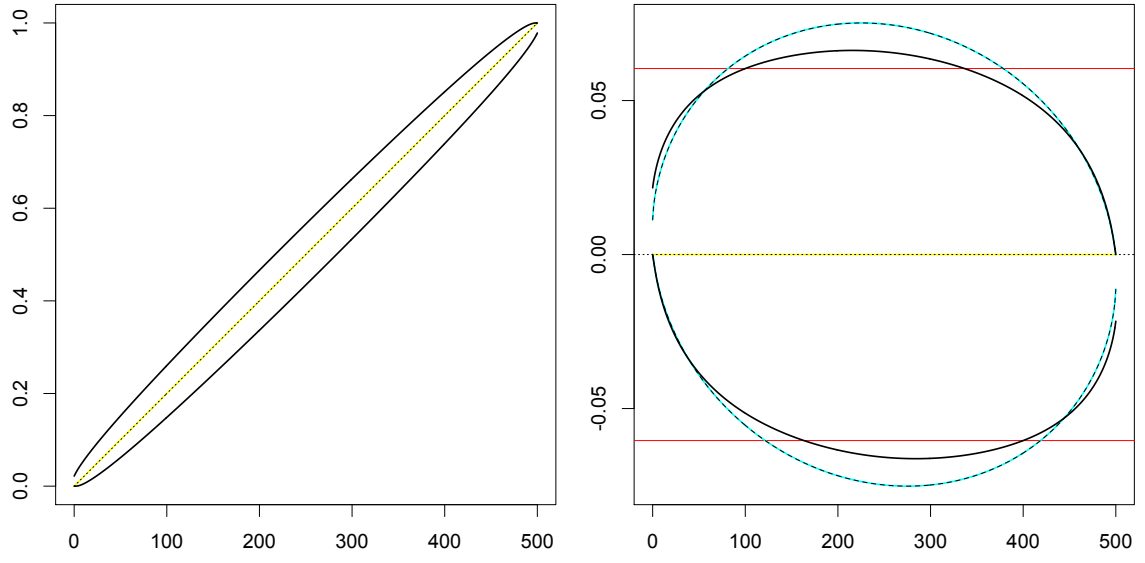


Figure 1: The confidence limits  $a_{nj}, b_{nj}$  (left panel) and the centered confidence limits  $a_{nj} - s_{nj}, b_{nj} - s_{nj}$  (right panel) for  $n = 500$ ,  $\nu = 1.1$  and  $\alpha = 5\%$ .

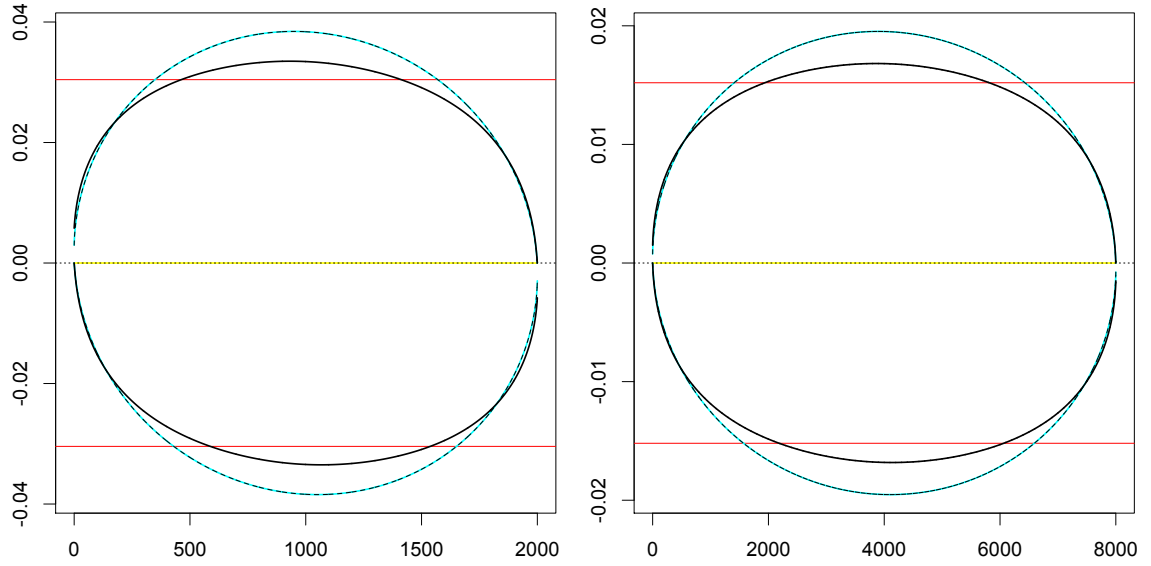


Figure 2: Centered confidence limits for  $n = 2000, 8000$  and  $\nu = 1.1$ ,  $\alpha = 5\%$ .

**An alternative approach via the union-intersection principle.** Aldor-Noiman et al. (2013) and Eiger et al. (2013) propose to use a union-intersection type goodness-of-fit test and related confidence bands. Under the null hypothesis that  $F \equiv F_o$ , the test statistic  $F_o(X_{n:i})$  and  $U_{n:i}$  follow a beta distribution with parameters  $i$  and  $n + 1 - i$ . Denoting its distribution function with  $B_{ni}$ , two resulting p-values would be  $B_{ni}(F_o(X_{n:i}))$  and  $1 - B_{ni}(F_o(X_{n:i}))$ . Thus one can reject the null hypothesis at level  $\alpha$  if the test statistic

$$\min_{i=1,2,\dots,n} \min\{B_{ni}(F_o(X_{n:i})), 1 - B_{ni}(F_o(X_{n:i}))\}$$

is lower or equal to the  $\alpha$ -quantile  $\kappa_{n,\alpha}^{\text{UI}}$  of

$$\min_{i=1,2,\dots,n} \min\{B_{ni}(U_{n:i}), 1 - B_{ni}(U_{n:i})\}. \quad (10)$$

A corresponding  $(1 - \alpha)$ -confidence band for  $F$  may be constructed as follows: With confidence  $1 - \alpha$  one may claim that for  $0 \leq j \leq n$  and  $X_{n:j} \leq x < X_{n:j+1}$ ,

$$F(x) \in [a_{nj}^{\text{UI}}, b_{nj}^{\text{UI}}],$$

where  $a_{n0}^{\text{UI}} := 0$ ,  $b_{nn}^{\text{UI}} := 1$ , and

$$\begin{aligned} a_{nj}^{\text{UI}} &:= B_{nj}^{-1}(\kappa_{n,\alpha}^{\text{UI}}) \quad \text{for } j > 1, \\ b_{nj}^{\text{UI}} &:= B_{nj}^{-1}(1 - \kappa_{n,\alpha}^{\text{UI}}) \quad \text{for } j < n. \end{aligned}$$

The results of Eiger et al. (2013) indicate that this goodness-of-fit test has similar properties as the one of Berk and Jones (1979). Indeed, if one considers the closely related test statistic

$$\max_{i=1,2,\dots,n} (n + 1)K(t_{ni}, F_o(X_{n:i}))$$

one may consider  $\exp(-(n + 1)K(t_{ni}, U_{ni}))$  as a simple surrogate for the minimum of the two p-values  $B_{ni}(F_o(X_{n:i}))$  and  $1 - B_{ni}(F_o(X_{n:i}))$ .

A possible weakness of the union-intersection approach is that it ignores correlations between the random variables  $U_{n:i}$ . Elementary calculations reveal that for  $1 \leq i < j \leq n$ ,

$$\text{Corr}(U_{n:i}, U_{n:j}) = \exp(-(\ell(t_{nj}) - \ell(t_{ni}))/2).$$

Thus the correlation of two neighbors  $U_{n:i}$  and  $U_{n:i+1}$  is rather large if  $t_{ni}$  is close to  $1/2$  but much smaller if  $t_{ni}$  is close to 0 or 1. As a result, the minimum in (10) tends to be attained for indices  $i$  such that  $t_{ni}$  is close to 0 or 1. With our additive correction term  $-C(t_{ni}) - \nu D(t_{ni})$  we try to account for such effects.

## 6 Proofs

### 6.1 Proofs for Section 2

**Proof of Theorem 2.2.** For symmetry reasons it suffices to prove upper bounds for

$$\mathbb{P}\left(\sup_{\mathcal{T} \cap [1/2, 1)} (X - C - \nu D) > \eta\right).$$

Note first that for  $t, t' \in (0, 1)$ ,

$$\left|\log \frac{t'(1-t')}{t(1-t)}\right| \leq \left|\log \frac{t'}{t}\right| + \left|\log \frac{1-t'}{1-t}\right| = |\text{logit}(t') - \text{logit}(t)|. \quad (11)$$

Consequently,

$$\begin{aligned} C(t') &= \log \log \left( \frac{e}{4t(1-t)} \frac{t(1-t)}{t'(1-t')} \right) \\ &\leq \log(\exp(C(t)) + |\text{logit}(t') - \text{logit}(t)|) \\ &= C(t) + \log(1 + \exp(-C(t)) |\text{logit}(t') - \text{logit}(t)|) \\ &\leq C(t) + |\text{logit}(t') - \text{logit}(t)| \end{aligned}$$

and since  $x \mapsto \log(1+x^2)$  has derivative  $2x/(1+x^2) \leq 1$ ,

$$D(t') \leq D(t) + |\text{logit}(t') - \text{logit}(t)|.$$

Now let  $(a_k)_{k \geq 0}$  be sequence of real numbers with  $a_0 = 0$  such that

$$a_k \rightarrow \infty \quad \text{and} \quad 0 < \delta_k := a_{k+1} - a_k \rightarrow 0 \quad \text{as } k \rightarrow \infty. \quad (12)$$

Then it follows from  $0 \leq \text{logit}(t) - \text{logit}(\ell(a_k)) \leq \delta_k$  for  $t \in [\ell(a_k), \ell(a_{k+1})]$  that

$$\begin{aligned} \sup_{\mathcal{T} \cap [\ell(a_k), \ell(a_{k+1})]} (X - C - \nu D) &\leq \sup_{\mathcal{T} \cap [\ell(a_k), \ell(a_{k+1})]} X - C(\ell(a_k)) - \nu D(\ell(a_k)) + (1 + \nu)\delta_k \\ &\leq \sup_{\mathcal{T} \cap [\ell(a_k), \ell(a_{k+1})]} X - C(\ell(a_k)) - \nu D(\ell(a_k)) + (1 + \nu)\delta_* \end{aligned}$$

with  $\delta_* := \max_{k \geq 0} \delta_k$ . Thus Condition 2.1 implies that

$$\begin{aligned} \mathbb{P}\left(\sup_{\mathcal{T} \cap [1/2, 1)} (X - C - \nu D) > \eta\right) &\leq \sum_{k \geq 0} \mathbb{P}\left(\sup_{\mathcal{T} \cap [\ell(a_k), \ell(a_{k+1})]} (X - C - \nu D) > \eta\right) \\ &\leq \sum_{k \geq 0} \mathbb{P}\left(\sup_{\mathcal{T} \cap [\ell(a_k), \ell(a_{k+1})]} X > \eta - (1 + \nu)\delta_* + C(\ell(a_k)) + \nu D(\ell(a_k))\right) \\ &\leq M \exp((1 + \nu)\delta_*) \exp(-\eta L(\delta_*)) \cdot G, \end{aligned}$$

where

$$\begin{aligned} G &:= \sum_{k \geq 0} \exp(-L(\delta_k)C(\ell(a_k)) - L(\delta_k)\nu D(\ell(a_k))) \\ &= \sum_{k \geq 0} \left( \log \frac{e}{4\ell'(a_k)} \right)^{-L(\delta_k)} \left( 1 + \left( \log \log \frac{e}{4\ell'(a_k)} \right)^2 \right)^{-\nu L(\delta_k)}. \end{aligned}$$

For any number  $a \geq 0$ ,

$$1 \leq \log \frac{e}{4\ell'(a)} = \log \frac{e(e^a + e^{-a} + 2)}{4} \in (a + \log(e/4), a + 1].$$

Now we define

$$a_k := \delta_* A(k) \quad \text{with} \quad A(s) := \frac{s}{\log(e + s)}$$

for some  $\delta_* > 0$  such that  $L(\delta_*) \geq L_o \in (0, 1)$ . Note that  $A(\cdot)$  is a continuously differentiable function on  $[0, \infty)$  with  $A(0) = 0$ , limit  $A(\infty) = \infty$  and derivative

$$A'(s) = \frac{1}{\log(e + s)} \left( 1 - \frac{s}{(e + s) \log(e + s)} \right) \in \left( 0, \frac{1}{\log(e + s)} \right).$$

This implies that (12) is indeed satisfied with

$$\delta_k \leq \frac{\delta_*}{\log(e + k)} = O((\log k)^{-1}) \quad \text{as } k \rightarrow \infty.$$

Moreover, as  $k \rightarrow \infty$ ,

$$\begin{aligned} & \left( \log \frac{e}{4\ell'(a_k)} \right)^{-L(\delta_k)} \left( 1 + \left( \log \log \frac{e}{4\ell'(a_k)} \right)^2 \right)^{-\nu L(\delta_k)} \\ &= O(a_k^{-L(\delta_k)} \log(a_k)^{-2\nu L(\delta_k)}) \\ &= O(k^{-L(\delta_k)} (\log k)^{L(\delta_k)} (\log k)^{-2\nu L(\delta_k)}) \\ &= O(k^{-1+O(1/\log k)} (\log k)^{-(2\nu-1)L(\delta_k)}) \\ &= O(k^{-1} (\log k)^{-(2\nu-1+o(1))}). \end{aligned}$$

Since  $2\nu - 1 > 1$ , this implies that  $G < \infty$ . Hence the asserted inequality is true with  $M_o = 2M \exp((1 + \nu)\delta_*) \cdot G$ .  $\square$

**Proof of Lemma 2.5.** To verify Condition 2.1 here, recall that if  $\mathbb{W} = (\mathbb{W}(t))_{t \geq 0}$  is standard Brownian motion, then  $(\mathbb{U}(t))_{t \in (0,1)}$  has the same distribution as  $((1 - t)\mathbb{W}(s(t)))_{t \in (0,1)}$  with

$s(t) := t/(1-t) = \exp(\text{logit}(t))$ . Hence for  $a \in \mathbb{R}$  and  $c \geq 0$ ,

$$\begin{aligned}
\sup_{t \in [\ell(a), \ell(a+c)]} X(t) &=_{\mathcal{L}} \sup_{t \in [\ell(a), \ell(a+c)]} \frac{(1-t)^2 \mathbb{W}(s(t))^2}{2t(1-t)} \\
&= \sup_{t \in [\ell(a), \ell(a+c)]} \frac{\mathbb{W}(s(t))^2}{2s(t)} \\
&= \sup_{s \in [e^a, e^{a+c}]} \frac{\mathbb{W}(s)^2}{2s} \\
&=_{\mathcal{L}} \sup_{u \in [1, e^c]} \frac{\mathbb{W}(u)^2}{2u}.
\end{aligned}$$

But it is well-known that  $(\mathbb{W}(u)/u)_{u \geq 1}$  is a reverse martingale. Thus  $(\exp(\lambda \mathbb{W}(u)/u))_{u \geq 1}$  is a nonnegative reverse submartingale for arbitrary real numbers  $\lambda$ . Hence it follows from Doob's inequality for nonnegative submartingales that for any  $\eta > 0$ ,

$$\begin{aligned}
\mathbb{P}\left(\sup_{u \in [1, e^c]} \frac{\mathbb{W}(u)^2}{2u} \geq \eta\right) &\leq \mathbb{P}\left(\sup_{u \in [1, e^c]} \frac{\mathbb{W}(u)^2}{u^2} \geq \frac{2\eta}{e^c}\right) \\
&\leq 2 \mathbb{P}\left(\sup_{u \in [1, e^c]} \mathbb{W}(u)/u \geq \sqrt{2e^{-c}\eta}\right) \\
&= 2 \inf_{\lambda > 0} \mathbb{P}\left(\sup_{u \in [1, e^c]} \exp(\lambda \mathbb{W}(u)/u) \geq \exp(\lambda \sqrt{2e^{-c}\eta})\right) \\
&\leq 2 \inf_{\lambda > 0} \mathbb{E} \exp(\lambda \mathbb{W}(1)/1) \exp(-\lambda \sqrt{2e^{-c}\eta}) \\
&= 2 \inf_{\lambda > 0} \exp(\lambda^2/2 - \lambda \sqrt{2e^{-c}\eta}) \\
&= 2 \exp(-e^{-c}\eta).
\end{aligned}$$

□

## 6.2 Various properties of the function $K(\cdot, \cdot)$

Before starting with a function  $K(\cdot, \cdot)$  itself, let us introduce two auxiliary functions:

$$\begin{aligned}
H(x) &:= x - \log(1+x), \quad x \in (-1, \infty), \\
\tilde{H}(z) &:= -\log(1-z) - z = H(-z), \quad z \in (-\infty, 1).
\end{aligned}$$

Elementary algebra shows that for  $s, t \in (0, 1)$ ,

$$K(s, t) = sH\left(\frac{t-s}{s}\right) + (1-s)\tilde{H}\left(\frac{t-s}{1-s}\right).$$

This representation will be useful for  $s$  close to 0 or 1.

**Lemma 6.1.** *Both functions  $H : [0, \infty) \rightarrow [0, \infty)$  and  $\tilde{H} : [0, 1) \rightarrow [0, \infty)$  are bijective, strictly increasing and strictly convex. Moreover,*

$$\begin{aligned} H(x) &\in \left[1 + x - \sqrt{1 + 2x}, \frac{x^2}{2 + x}\right] \quad \text{for } x \in [0, \infty), \\ \tilde{H}(z) &\in [-\log(1 - z^2)/2, -\log(1 - z)] \quad \text{for } z \in [0, 1). \end{aligned}$$

*The inverse functions  $H^{-1} : [0, \infty) \rightarrow [0, \infty)$  and  $\tilde{H}^{-1} : [0, \infty) \rightarrow [0, 1)$  are strictly increasing and strictly concave with*

$$\begin{aligned} H^{-1}(y) &\in [\sqrt{2y + y^2/4} + y/2, \sqrt{2y + y}], \\ \tilde{H}^{-1}(y) &\in [1 - e^{-y}, \sqrt{1 - e^{-2y}}]. \end{aligned}$$

The proof of this lemma is elementary and thus omitted. Now we are ready to state essential properties of  $K(\cdot, \cdot)$ :

**(K.0)** With the convention that  $0 \log 0 := 0$  one can easily verify that the function  $K : [0, 1] \times (0, 1) \rightarrow \mathbb{R}$  is continuous. In particular,  $K(0, t) = -\log(1 - t)$  and  $K(1, t) = -\log t$ . Moreover,  $K(1 - s, 1 - t) = K(s, t)$  for arbitrary  $s \in [0, 1]$  and  $t \in (0, 1)$ .

**(K.1)** For  $s, t \in (0, 1)$ ,

$$\frac{\partial K(s, t)}{\partial s} = \text{logit}(s) - \text{logit}(t) \quad \text{and} \quad \frac{\partial K(s, t)}{\partial t} = -\frac{s}{t} + \frac{1 - s}{1 - t} = \frac{t - s}{t(1 - t)}.$$

(The latter formula is true even for  $s \in [0, 1]$ .) In particular,  $K(s, t) \geq 0$  with equality if, and only if,  $s = t$ .

**(K.2)** For  $s, t \in (0, 1)$ ,

$$\begin{aligned} \frac{\partial^2 K(s, t)}{\partial s^2} &= \frac{1}{s(1 - s)}, \quad \frac{\partial^2 K(s, t)}{\partial s \partial t} = -\frac{1}{t(1 - t)} \quad \text{and} \\ \frac{\partial^2 K(s, t)}{\partial t^2} &= \frac{s}{t^2} + \frac{1 - s}{(1 - t)^2} = \frac{(t - s)^2 + s(1 - s)}{t^2(1 - t)^2}. \end{aligned}$$

In particular, the Hessian matrix of  $K$  at  $(s, t)$  has positive diagonal elements and non-negative determinant  $(t - s)^2 / (s(1 - s)t^2(1 - t)^2)$ . This implies that  $K$  is convex on  $[0, 1] \times (0, 1)$ .



**(K.3)** For fixed  $u \in (0, 1)$  and arbitrary  $0 < t < t' < 1$ ,

$$\frac{K(0, t')}{K(0, t)}, \frac{K(t'u, t')}{K(tu, t)}, \frac{K(t', t'u)}{K(t, tu)} \in \left( \frac{t'}{t}, \frac{t'(1-t)}{(1-t')t} \right).$$

*Proof.* Since  $K(tu, tu) = 0$ , it follows from (K.1) that

$$K(tu, t) = \int_{tu}^t \frac{\partial K(tu, x)}{\partial x} dx = \int_{tu}^t \frac{(x - tu)}{x(1-x)} dx = \int_u^1 \frac{t(v-u)}{v(1-tv)} dv.$$

These formulae remain true if we replace  $u$  with 0. On the other hand, since  $K(tu, tu) = 0 = \partial K(s, tu)/\partial s$  for  $s = tu$ , a suitable version of Taylor's formula and (K.2) imply that

$$K(t, tu) = \int_{tu}^t (t-x) \frac{\partial^2}{\partial x^2} K(x, tu) dx = \int_{tu}^t \frac{(t-x)}{x(1-x)} dx = \int_u^1 \frac{t(1-v)}{v(1-tv)} dv,$$

But for any  $v \in (0, 1)$ ,

$$\frac{\partial}{\partial t} \log \frac{t}{1-tv} = \frac{1}{t(1-tv)} \in \left( \frac{1}{t}, \frac{1}{t(1-t)} \right) = (\log'(t), \text{logit}'(t)).$$

Thus for  $0 < t < t' < 1$ ,

$$\frac{t'}{1-t'v} / \frac{t}{1-tv} \in \left( \frac{t'}{t}, \frac{t'(1-t)}{(1-t')t} \right),$$

and this entails the asserted inequalities for the three ratios  $K(0, t')/K(0, t)$ ,  $K(t'u, t')/K(tu, t)$  and  $K(t', t'u)/K(t, tu)$ .  $\square$

**(K.4)** To verify Theorems 3.2, 3.4 and 5.1 we have to approximate  $K$  by a simpler function  $\tilde{K}$  given by

$$\tilde{K}(s, t) := \frac{(s-t)^2}{2t(1-t)}.$$

Indeed, for arbitrary  $s, t \in (0, 1)$  and  $c := |\text{logit}(s) - \text{logit}(t)|$ ,

$$\frac{K(s, t)}{\tilde{K}(s, t)}, \frac{K(s, t)}{\tilde{K}(t, s)} \in [e^{-c}, e^c].$$

*Proof.* It follows from (K.1-2) and Taylor's formula that

$$K(s, t) = \frac{(s-t)^2}{2\xi(1-\xi)}$$

for some  $\xi$  between  $\min\{s, t\}$  and  $\max\{s, t\}$ . Hence

$$\frac{K(s, t)}{\tilde{K}(s, t)} = \frac{t(1-t)}{\xi(1-\xi)} \quad \text{and} \quad \frac{K(s, t)}{\tilde{K}(t, s)} = \frac{s(1-s)}{\xi(1-\xi)}$$

are both contained in  $[e^{-c}, e^c]$ , according to (11).  $\square$

**(K.5)** For arbitrary  $\gamma > 0$  and  $s \in [0, 1], t \in (0, 1)$ , the inequality  $K(s, t) \leq \gamma$  implies that

$$(t - s)^\pm \leq \begin{cases} \sqrt{2\gamma s(1-s)} + (1-2s)^\pm \gamma, \\ \sqrt{2\gamma t(1-t)} + (2t-1)^\pm \gamma. \end{cases}$$

In particular,

$$|s - t| \leq \min\{\sqrt{2s(1-s)\gamma}, \sqrt{2t(1-t)\gamma}\} + \gamma.$$

*Proof.* The first inequality has been proved by Dümbgen (1998), but for the reader's convenience and the proof of the new part, a complete derivation is given here: For symmetry reasons, it suffices to consider the case  $0 \leq s < t < 1$  and derive the upper bounds for  $\delta := t - s = (t - s)^+$ .

Let us first treat the case  $s = 0$ : Here  $K(s, t) = -\log(1 - t) \geq t$ . Thus  $K(0, t) \leq \gamma$  implies that  $\delta = t \leq \gamma = \sqrt{2\gamma s(1-s)} + (1-2s)\gamma$ . Moreover,  $\sqrt{2\gamma t(1-t)} + (2t-1)^+\gamma \geq t(\sqrt{2(1-t)} + (2t-1)^+)$ , and elementary considerations show that  $\sqrt{2(1-t)} + (2t-1)^+ \geq 1$ .

Now let  $0 < s < t < 1$  and  $\delta := t - s$ . It follows from  $K(s, s) = 0$  and (K.1) that

$$K(s, t) = \int_s^t \frac{\partial K(s, y)}{\partial y} dy = \int_0^\delta \frac{x}{(s+x)(1-s-x)} dx \geq \int_0^\delta \frac{x}{s(1-s) + (1-2s)x} dx.$$

In case of  $s \geq 1/2$ , the latter integral is not smaller than  $\delta^2/(2s(1-s))$ , and  $K(s, t) \leq \gamma$  implies the upper bound  $\delta \leq \sqrt{2\gamma s(1-s)}$ . In case of  $s < 1/2$ , we obtain the bound

$$K(s, t) \geq \int_0^\delta \frac{x}{\alpha + \beta x} dx = \frac{\delta}{\beta} - \frac{\alpha}{\beta^2} \log\left(1 + \frac{\beta\delta}{\alpha}\right) = \frac{\alpha}{\beta^2} H\left(\frac{\beta\delta}{\alpha}\right)$$

with  $\alpha := s(1-s) > 0$ ,  $\beta := 1-2s > 0$  and the auxiliary function  $H$  from Lemma 6.1.

Consequently, the inequality  $K(s, t) \leq \gamma$  entails that  $H(\beta\delta/\alpha) \leq \beta^2\gamma/\alpha$ , so

$$\delta \leq (\alpha/\beta)H^{-1}(\beta^2\gamma/\alpha) \leq \sqrt{2\gamma\alpha} + \beta\gamma = \sqrt{2\gamma s(1-s)} + (1-2s)\gamma.$$

On the other hand,

$$K(s, t) = \int_s^t \frac{y-s}{y(1-y)} dy = \int_0^\delta \frac{\delta-x}{(t-x)(1-t+x)} dx \geq \int_0^\delta \frac{\delta-x}{t(1-t) + (2t-1)x} dx.$$

In case of  $t \leq 1/2$ , the latter integral is at least  $\delta^2/(2t(1-t))$ , and we may conclude from  $K(s, t) \leq \gamma$  that  $\delta$  is bounded by  $\sqrt{2\gamma t(1-t)}$ . In case of  $t > 1/2$ , we define  $a := t(1-t) > 0$ ,  $b := 2t-1 > 0$  and may write

$$K(s, t) \geq \int_0^\delta \frac{\delta-x}{a+bx} dx > \int_0^\delta \frac{x}{a+bx} dx = \frac{a}{b^2} H\left(\frac{b\delta}{a}\right).$$

The second inequality in the previous display follows from the fact that  $f(x) := 1/(a + bx)$  is strictly decreasing on  $[0, \delta]$ . Thus  $\int_0^\delta (\delta - x)f(x) dx - \int_0^\delta xf(x) dx$  equals

$$\int_0^\delta (\delta - 2x)f(x) dx = \int_0^\delta (\delta - 2x)(f(x) - f(\delta/2)) dx$$

and is strictly positive. Hence the preceding considerations yield the upper bound  $\sqrt{2\gamma a} + b\gamma = \sqrt{2\gamma t(1-t)} + (2t-1)\gamma$  for  $\delta$ .  $\square$

**(K.6)** For  $s \in (0, 1)$  and  $\gamma > 0$  let  $b = b(s, \gamma) \in (s, 1)$  solve the equation

$$K(s, b) = \gamma.$$

Then

$$\frac{b-s}{sH^{-1}(\gamma/s)} \begin{cases} \leq 1, \\ \rightarrow 1 \end{cases} \text{ as } s, \gamma \rightarrow 0, \quad (13)$$

$$\frac{b-s}{\sqrt{2\gamma s(1-s)}} \rightarrow 1 \text{ as } \frac{\gamma}{s(1-s)} \rightarrow 0, \quad (14)$$

$$\frac{b-s}{(1-s)\tilde{H}^{-1}(\gamma/(1-s))} \in [s, 1]. \quad (15)$$

*Proof.* With  $\delta := (b-s)/s > 0$  we may write

$$\frac{\gamma}{s} = \frac{K(s, s+s\delta)}{s} = H(\delta) + \frac{1-s}{s}\tilde{H}\left(\frac{s\delta}{1-s}\right).$$

Since  $\tilde{H} \geq 0$ , this implies that  $H(\delta) \leq \gamma/s$ , which is equivalent to  $b-s \leq sH^{-1}(\gamma/s)$ . On the other hand, it follows from the expansion  $-\log(1-z) = \sum_{k=1}^\infty z^k/k = z + \tilde{H}(z)$  that

$$\frac{\gamma}{s} = H(\delta) + \frac{1-s}{s} \sum_{k=2}^\infty \left(\frac{s\delta}{1-s}\right)^k / k \leq H(\delta) + \frac{s\delta^2}{2(1-s-s\delta)}.$$

As  $c := \max\{s, \gamma\} \rightarrow 0$ , it follows from  $\delta \leq H^{-1}(\gamma/s) \leq \sqrt{2\gamma/s} + \gamma/s$  that

$$\begin{aligned} 1-s-s\delta &\geq 1-s-\sqrt{2s\gamma}-\gamma = 1-O(c), \\ s\delta^2 &\leq s(\sqrt{2\gamma/s} + \gamma/s)^2 = O(c)\gamma/s, \end{aligned}$$

whence

$$\frac{\gamma}{s} \leq H(\delta) + O(c)\frac{\gamma}{s}.$$

Consequently,

$$b-s \geq sH^{-1}((1-O(c))\gamma/s) \geq (1-O(c))sH^{-1}(\gamma/s),$$

the latter inequality following from concavity of  $H^{-1}$ . This proves (13).

As to (14), let  $c := \sqrt{\gamma/(s(1-s))} < 1/2$ , and define the points  $t(x) = t(s, \gamma, x) := s + \sqrt{2\gamma s(1-s)}x = s + cs(1-s)\sqrt{2x}$  for  $x \in [0, 2]$ . Then

$$0 < \text{logit}(t(x)) - \text{logit}(s) < \log \frac{1+2c}{1-2c} \rightarrow 0 \quad \text{as } c \rightarrow 0.$$

Consequently, by (K.4),

$$K(s, t(x)) = (1 + o(1))\tilde{K}(t(x), s) = (1 + o(1))\gamma x$$

uniformly in  $x \in [0, 2]$ . This shows that  $b(s, \gamma) = t(1 + o(1)) = s + \sqrt{2\gamma s(1-s)}(1 + o(1))$  as  $c \rightarrow 0$ .

Finally, let  $\delta := (b - s)/(1 - s)$ . Then it follows from  $\tilde{H}(z) \geq z^2/2 \geq H(z)$  that

$$\frac{\gamma}{1-s} = \frac{s}{1-s}H\left(\frac{1-s}{s}\delta\right) + \tilde{H}(\delta) \begin{cases} \geq \tilde{H}(\delta), \\ \leq (1-s)\delta^2/(2s) + \tilde{H}(\delta) \leq \tilde{H}(\delta)/s. \end{cases}$$

Consequently, by concavity of  $\tilde{H}^{-1}(\cdot)$ ,

$$s\tilde{H}^{-1}(\gamma/(1-s)) \leq \tilde{H}^{-1}(s\gamma/(1-s)) \leq \delta \leq \tilde{H}^{-1}(\gamma/(1-s)),$$

which yields (15). □

### 6.3 Proofs for Section 3

Before proving Lemma 3.1 let us recall that for  $s \in \mathbb{R}$  and  $t \in (0, 1)$ ,

$$K(s, t) := \sup_{\lambda \in \mathbb{R}} (\lambda s - \log(1 - t + te^\lambda)) = \begin{cases} s \log \frac{s}{t} + (1-s) \log \frac{1-s}{1-t} & \text{if } s \in [0, 1], \\ \infty & \text{else.} \end{cases}$$

Indeed, Hoeffding (1963) showed that for a random variable  $Y \sim \text{Bin}(n, t)$  and  $s \in \mathbb{R}$ ,

$$\begin{aligned} \mathbb{P}(Y \geq ns) &\leq \exp\left(-n \sup_{\lambda \geq 0} (\lambda s - \log(1 - t + te^\lambda))\right) = \exp(-nK(s, t)) \quad \text{if } s \geq t, \\ \mathbb{P}(Y \leq ns) &\leq \exp\left(-n \sup_{\lambda \leq 0} (\lambda s - \log(1 - t + te^\lambda))\right) = \exp(-nK(s, t)) \quad \text{if } s \leq t. \end{aligned}$$

**Proof of Lemma 3.1.** We imitate and modify a martingale argument of Berk and Jones (1979, Lemma 4.3) which goes back to Kiefer (1973). Note first that  $\hat{G}_n(t)/t$  is a reverse martingale in  $t \in (0, 1)$ , that means,

$$\mathbb{E}(\hat{G}_n(s)/s \mid (\hat{G}_n(t'))_{t' \geq t}) = \hat{G}_n(t)/t \quad \text{for } 0 < s < t < 1.$$

Consequently, for  $0 < t < t' < 1$  and  $0 \leq u \leq 1$ ,

$$\begin{aligned} \mathbb{P}\left(\inf_{s \in [t, t']} \widehat{G}_n(s)/s \leq u\right) &= \inf_{\lambda \leq 0} \mathbb{P}\left(\sup_{s \in [t, t']} \exp(\lambda \widehat{G}_n(s)/s - \lambda u) \geq 1\right) \\ &\leq \inf_{\lambda \leq 0} \mathbb{E} \exp(\lambda \widehat{G}_n(t)/t - \lambda u) \end{aligned}$$

by Doob's inequality for non-negative submartingales. But  $n\widehat{G}_n(t) \sim \text{Bin}(n, t)$ , so

$$\begin{aligned} \inf_{\lambda \leq 0} \mathbb{E} \exp(\lambda \widehat{G}_n(t)/t - \lambda u) &= \inf_{\lambda \leq 0} \mathbb{E} \exp(\lambda n \widehat{G}(t) - n \lambda t u) \\ &= \exp\left(-n \sup_{\lambda \leq 0} (\lambda t u - \log(1 - t + t e^\lambda))\right) \\ &= \exp(-n K(tu, t)). \end{aligned}$$

Thus

$$\mathbb{P}\left(\inf_{s \in [t, t']} \widehat{G}_n(s)/s \leq u\right) \leq \exp(-n K(tu, t)) \quad \text{for all } u \in [0, 1].$$

One may rewrite this inequality as

$$\mathbb{P}\left(\sup_{s \in [t, t']} nK(t \min\{\widehat{G}_n(s)/s, 1\}, t) \geq \eta\right) \leq \exp(-\eta) \quad \text{for all } \eta \geq 0.$$

For if  $\eta > -n \log(1 - t)$ , the probability on the left hand side equals 0. Otherwise there exists a unique  $u = u(t, \eta) \in [0, 1]$  such that  $nK(tu, t) = \eta$ . But then

$$nK(t \min\{\widehat{G}_n(s)/s, 1\}, t) \geq \eta \quad \text{if, and only if,} \quad \widehat{G}_n(s)/s \leq u.$$

Finally, it follows from property (K.3) of  $K(\cdot, \cdot)$  that for  $t \leq s \leq t'$ ,

$$K(\min\{\widehat{G}_n(s), s\}, s) = K(s \min\{\widehat{G}_n(s)/s, 1\}, s) \leq e^c K(t \min\{\widehat{G}_n(s)/s, 1\}, t)$$

with  $c := \text{logit}(t') - \text{logit}(t)$ . Hence

$$\mathbb{P}\left(\sup_{s \in [t, t']} nK(\min\{\widehat{G}_n(s), s\}, s) \geq \eta\right) \leq \exp(-e^{-c}\eta) \quad \text{for all } \eta \geq 0.$$

Since  $(\widehat{G}_n(t))_{t \in (0,1)}$  has the same distribution as  $(1 - \widehat{G}_n((1 - t) -))_{t \in (0,1)}$ , and because of the symmetry relations  $K(s, t) = K(1 - s, 1 - t)$  and  $\text{logit}(1 - t) = -\text{logit}(t)$ , the previous inequality implies further that

$$\begin{aligned} &\mathbb{P}\left(\sup_{s \in [t, t']} nK(\max\{\widehat{G}_n(s), s\}, s) \geq \eta\right) \\ &= \mathbb{P}\left(\sup_{s \in [t, t']} nK(\min\{1 - \widehat{G}_n(s), 1 - s\}, 1 - s) \geq \eta\right) \\ &= \mathbb{P}\left(\sup_{s \in [1-t', 1-t]} nK(\min\{\widehat{G}_n(s), s\}, s) \geq \eta\right) \\ &\leq \exp(-e^{-c}\eta) \quad \text{for all } \eta \geq 0. \end{aligned}$$

Consequently, since  $K(\cdot, s) = \max\{K(\min\{\cdot, s\}, s), K(\max\{\cdot, s\}, s)\}$ ,

$$\mathbb{P}\left(\sup_{s \in [t, t']} nK(\widehat{G}_n(s), s) \geq \eta\right) \leq 2 \exp(-e^{-c}\eta) \quad \text{for all } \eta \geq 0.$$

□

**Proof of Theorem 3.2.** For any fixed  $\delta \in (0, 1/2)$ , it follows from Donsker's invariance principle for the uniform empirical process and the continuous mapping theorem that

$$\sup_{t \in [-\delta, \delta]} \left( \frac{\mathbb{U}_n(t)^2}{2t(1-t)} - C(t) - \nu D(t) \right) \rightarrow_{\mathcal{L}} \sup_{[-\delta, \delta]} (X - C - \nu D),$$

where  $X(t) = \mathbb{U}(t)^2/(2t(1-t))$ . With  $X_n(t) = nK(\widehat{G}_n(t), t)$  it follows from property (K.4) of  $K(\cdot, \cdot)$  that

$$\frac{\mathbb{U}_n(t)^2}{2t(1-t)} = n\tilde{K}(\widehat{G}_n(t), t) = X_n(t)(1 + r_n(t))$$

with

$$\sup_{t \in [\delta, 1-\delta]} |r_n(t)| \leq (1 - n^{-1/2}\delta^{-1}\|\mathbb{U}_n\|_\infty)^{-2} - 1 = O_p(n^{-1/2}).$$

Thus

$$\sup_{[-\delta, \delta]} (X_n - C - \nu D) \rightarrow_{\mathcal{L}} \sup_{[-\delta, \delta]} (X - C - \nu D).$$

But Theorem 2.2 implies that for any  $1 < \nu' < \nu$ , the random variables  $T_{n, \nu'}$  and  $T_{\nu'}$  satisfy the inequalities  $\mathbb{P}(T_{n, \nu'} > \eta) \leq M_o \exp(-L_o \eta)$  and  $\mathbb{P}(T_{\nu'} > \eta) \leq M_o \exp(-L_o \eta)$  for arbitrary  $\eta \in \mathbb{R}$  and some constants  $L_o \in (0, 1)$ ,  $M_o \geq 1$ . Consequently for any  $\rho > 0$ ,

$$\begin{aligned} \mathbb{P}\left(\sup_{[\delta, 1-\delta]} (X_n - C - \nu D) < \sup_{(0,1)} (X_n - C - \nu D)\right) \\ \leq \mathbb{P}(T_{n, \nu'} - (\nu - \nu')D(\delta) > -\rho) + \mathbb{P}(X_n(1/2) \leq -\rho) \\ \leq M_o \exp(L_o \rho - L_o(\nu - \nu')D(\delta)) + \mathbb{P}(X(1/2) \leq -\rho) + o(1) \end{aligned}$$

because  $X_n(1/2) \rightarrow_{\mathcal{L}} X(1/2)$ , and

$$\begin{aligned} \mathbb{P}\left(\sup_{[\delta, 1-\delta]} (X - C - \nu D) < \sup_{(0,1)} (X - C - \nu D)\right) \\ \leq M_o \exp(L_o \rho - L_o(\nu - \nu')D(\delta)) + \mathbb{P}(X(1/2) \leq -\rho). \end{aligned}$$

Setting  $\rho = (\nu - \nu')D(\delta)/2$ , the limits of the right hand sides become arbitrarily small for sufficiently small  $\delta$ . This shows that  $T_{n, \nu} = \sup_{(0,1)} (X_n - C - \nu D)$  converges in distribution to  $T_\nu$ . □

Our proof of Lemma 3.3 involves an exponential inequality for Beta distributions from Dümbgen (1998). For the reader's convenience, its proof is included in the supplementary material.

**Lemma 6.2.** *Let  $s, t \in (0, 1)$ , and let  $Y \sim \text{Beta}(mt, m(1-t))$  for some  $m > 0$ . Then*

$$\begin{aligned}\mathbb{P}(Y \leq s) &\leq \inf_{\lambda \leq 0} \mathbb{E} \exp(\lambda Y - \lambda s) \leq \exp(-mK(t, s)) \quad \text{if } s \leq t, \\ \mathbb{P}(Y \geq s) &\leq \inf_{\lambda \geq 0} \mathbb{E} \exp(\lambda Y - \lambda s) \leq \exp(-mK(t, s)) \quad \text{if } s \geq t.\end{aligned}$$

**Proof of Lemma 3.3.** We utilize a well-known representation of uniform order statistics: Let  $E_1, E_2, \dots, E_{n+1}$  be independent random variables with standard exponential distribution, i.e. Gamma(1), and let  $S_j := \sum_{i=1}^j E_i$ . Then

$$(U_{ni})_{i=1}^n =_{\mathcal{L}} (S_i/S_{n+1})_{i=1}^n.$$

In particular,  $U_{n:i} \sim \text{Beta}(i, n+1-i) = \text{Beta}((n+1)t_{ni}, (n+1)(1-t_{ni}))$  and  $\mathbb{E} U_{n:i} = t_{ni}$ . Furthermore, for  $2 \leq k \leq n+1$ , the random vectors  $(S_i/S_k)_{i=1}^{k-1}$  and  $(S_i)_{i=k}^{n+1}$  are stochastically independent. This implies that  $(U_{n:i}/t_{ni})_{i=1}^n$  is a reverse martingale, because for  $1 \leq j < k \leq n$ ,

$$\mathbb{E}\left(\frac{U_{n:j}}{t_{nj}} \mid (S_i)_{i=k}^{n+1}\right) = \mathbb{E}\left(\frac{S_j}{t_{nj}S_k} \cdot \frac{S_k}{S_{n+1}} \mid (S_i)_{i=k}^{n+1}\right) = \frac{j}{t_{nj}k} \cdot \frac{S_k}{S_{n+1}} = \frac{U_{n:k}}{t_{nk}}.$$

Consequently, for  $1 \leq j \leq k \leq n$  and  $0 < u < 1$ , it follows from Doob's inequality and Lemma 6.2 that

$$\begin{aligned}\mathbb{P}\left(\min_{j \leq i \leq k} \frac{U_{n:i}}{t_{ni}} \leq u\right) &= \inf_{\lambda < 0} \mathbb{P}\left(\min_{j \leq i \leq k} \exp\left(\lambda \frac{U_{n:i}}{t_{ni}} - \lambda u\right) \geq 1\right) \\ &\leq \inf_{\lambda < 0} \mathbb{E} \exp(\lambda U_{n:j} - \lambda u t_{nj}) \\ &\leq \exp(-(n+1)K(t_{nj}, t_{nj}u)).\end{aligned}$$

Again one may reformulate the previous inequalities as follows: For any  $\eta > 0$ ,

$$\mathbb{P}\left(\max_{j \leq i \leq k} (n+1)K\left(t_{nj}, t_{nj} \min\left\{\frac{U_{n:i}}{t_{ni}}, 1\right\}\right) \geq \eta\right) \leq \exp(-\eta).$$

But property (K.3) of  $K(\cdot, \cdot)$  implies that for  $j \leq i \leq k$ ,

$$K(t_{ni}, \min\{U_{n:i}, t_{ni}\}) \leq e^c K\left(t_{nj}, t_{nj} \min\left\{\frac{U_{n:i}}{t_{ni}}, 1\right\}\right)$$

with  $c := \text{logit}(t_{nk}) - \text{logit}(t_{nj})$ . Consequently,

$$\mathbb{P}\left(\max_{j \leq i \leq k} (n+1)K(t_{ni}, \min\{U_{n:i}, t_{ni}\}) \geq \eta\right) \leq \exp(-e^{-c}\eta) \quad \text{for all } \eta > 0.$$

Since  $(1 - U_{n:n+1-i})_{i=1}^n$  has the same distribution as  $(U_{n:i})_{i=1}^n$ , a symmetry argument as in the proof of Lemma 3.1 reveals that

$$\mathbb{P}\left(\max_{j \leq i \leq k} (n+1)K(t_{ni}, U_{n:i}) \geq \eta\right) \leq 2 \exp(-e^{-c}\eta) \quad \text{for all } \eta > 0.$$

□

**Proof of Theorem 3.4.** One can use essentially the same arguments as in the proof of Theorem 3.2. This time one has to utilize the well-known fact that

$$(U_{n:i})_{i=1}^n = (t_{ni} + n^{-1/2}\mathbb{V}_n(t_{ni}))_{i=1}^n$$

where the uniform quantile process  $\mathbb{V}_n$  with  $\mathbb{V}_n(t) := \sqrt{n}(\widehat{G}_n^{-1}(t) - t)$  converges in distribution in  $\ell_\infty([0, 1])$  to a Brownian bridge  $\mathbb{V}$ ; see e.g. Shorack and Wellner (1986), pages 86, 93, and 637-644. □

## 6.4 Proofs for Sections 4 and 5

**Proof of Lemma 4.1.** Suppose that  $T_{n,\nu}(F_o) \leq \kappa$ . Then the inequalities in (K.5) imply that

$$|\widehat{F}_n - F_o| \leq \sqrt{2\tilde{\Gamma}(F_o)F_o(1-F_o)/n + \tilde{\Gamma}(F_o)/n},$$

where  $\tilde{\Gamma}(t) := C(t) + \nu D(t) + \kappa$ . Multiplying this inequality with  $n$  and utilizing the triangle inequality  $|\widehat{F}_n - F_o| \geq |F - F_o| - |\widehat{F}_n - F|$  leads to

$$n|F - F_o| \leq \sqrt{2n\tilde{\Gamma}(F_o)F_o(1-F_o) + \tilde{\Gamma}(F_o)} + n|\widehat{F}_n - F|. \quad (16)$$

Now our goal is to get rid of the term  $n|\widehat{F}_n - F|$  on the right hand side. Defining the auxiliary stochastic process

$$W_n := \frac{n(\widehat{F}_n - F)^2}{F(1-F)}$$

with the convention  $0/0 := 0$ , we may rewrite (16) as

$$\begin{aligned} n|F - F_o| &\leq \sqrt{2n\tilde{\Gamma}(F_o)F_o(1-F_o) + \tilde{\Gamma}(F_o)} + \sqrt{W_n n F(1-F)} \\ &\leq \sqrt{2n\tilde{\Gamma}(F_o)F_o(1-F_o) + \tilde{\Gamma}(F_o)} + \sqrt{W_n n F_o(1-F_o)} + \sqrt{W_n n |F - F_o|} \\ &\leq \sqrt{n(4\tilde{\Gamma}(F_o) + 2W_n)F_o(1-F_o) + \tilde{\Gamma}(F_o)} + \sqrt{W_n n |F - F_o|}, \end{aligned} \quad (17)$$

where we utilized the inequalities  $|a(1-a) - b(1-b)| \leq |a-b|$  for  $a, b \in [0, 1]$  and  $\sqrt{c+d} \leq \sqrt{c} + \sqrt{d} \leq \sqrt{2c+2d}$  for  $c, d \geq 0$ . Note that inequality (17) is of the form  $Y_n \leq V_n + \sqrt{W_n Y_n}$  with the



nonnegative processes  $Y_n = n|F - F_o|$  and  $V_n = \sqrt{n(4\tilde{\Gamma}(F_o) + 2W_n)F_o(1 - F_o) + \tilde{\Gamma}(F_o)}$ . But  $Y_n \leq V_n + \sqrt{W_n Y_n}$  is equivalent to  $Y_n/V_n \leq 1 + \sqrt{W_n/V_n} \sqrt{Y_n/V_n}$ , and this may be rewritten as  $(\sqrt{Y_n/V_n} - \sqrt{W_n/V_n}/2)^2 \leq 1 + (W_n/V_n)/4$ , so

$$\sqrt{Y_n/V_n} \leq \sqrt{W_n/V_n}/2 + \sqrt{1 + (W_n/V_n)/4} \leq 1 + \sqrt{W_n/V_n}.$$

Consequently,

$$\begin{aligned} n|F - F_o| &\leq (1 + \sqrt{W_n/V_n})^2 \left( \sqrt{n(4\tilde{\Gamma}(F_o) + 2W_n)F_o(1 - F_o) + \tilde{\Gamma}(F_o)} \right) \\ &\leq (1 + \sqrt{W_n/\kappa})^2 \sqrt{4 + 2W_n/\kappa} \left( \sqrt{n\tilde{\Gamma}(F_o)F_o(1 - F_o) + \tilde{\Gamma}(F_o)} \right) \\ &\leq 2(1 + \sqrt{W_n/\kappa})^{5/2} \left( \sqrt{n\tilde{\Gamma}(F_o)F_o(1 - F_o) + \tilde{\Gamma}(F_o)} \right), \end{aligned}$$

because  $V_n \geq \tilde{\Gamma}(F_o) \geq \kappa$ . Finally, since  $B' = B'_{\nu, \kappa} := \max\{\sup_{(0,1)} \tilde{\Gamma}/\Gamma, 1\} < \infty$ , we obtain the inequality

$$\frac{n|F - F_o|}{\sqrt{n\Gamma(F_o)F_o(1 - F_o)} + \Gamma(F_o)} \leq 2B'(1 + \sqrt{W_n/\kappa})^{5/2} \quad \text{if } T_{n,\nu}(F_o) \leq \kappa. \quad (18)$$

On the left hand side stands a function  $\Delta_n = \Delta_n(\cdot, F, F_o)$ , and its supremum over  $\mathbb{R}$  equals  $\Delta_n(F, F_o)$ . Thus it suffices to show that for a suitable constant  $B_{\nu, \kappa}$ ,

$$\mathbb{P}_F \left( 2B'(1 + \sqrt{W_n(x)/\kappa})^{5/2} \geq \Delta_n(x) \right) \leq B_{\nu, \kappa} \Delta_n(x)^{-4/5}$$

for any  $x \in \mathbb{R}$ . Indeed,  $2B'(1 + \sqrt{W_n(x)/\kappa})^{5/2} \geq \Delta_n(x)$  is equivalent to

$$W_n(x) \geq \kappa \max\{0, \Delta_n(x)^{2/5} (2B')^{-2/5} - 1\}^2.$$

Since  $\mathbb{E} W_n(x) \leq 1$ , it follows from Markov's inequality that the latter inequality occurs with probability at most

$$\kappa^{-1} \max\{0, \Delta_n(x)^{2/5} (2B')^{-2/5} - 1\}^{-2} = \max\{0, B'' \Delta_n(x)^{2/5} - \kappa^{1/2}\}^{-2}$$

with a certain constant  $B'' = B''_{\nu, \kappa}$ . This bound is trivial if  $B'' \Delta_n(x)^{2/5} < 1 + \kappa^{1/2}$ , which is equivalent to  $\Delta_n(x)^{4/5} < B_{\nu, \kappa} := (1 + \kappa^{1/2})^2 / (B'')^2$ . Otherwise,

$$\begin{aligned} &\max\{0, B'' \Delta_n(x)^{2/5} - \kappa^{1/2}\}^{-2} \\ &= (B'' - \kappa^{1/2} \Delta_n(x)^{-2/5})^{-2} \Delta_n(x)^{-4/5} \leq B \Delta_n(x)^{-4/5}. \end{aligned}$$

□

**Proof of Lemma 4.2.** In what follows we use frequently the elementary inequalities

$$\frac{\phi(x)}{x+1} \leq \Phi(-x) \leq \frac{\phi(x)}{x} \quad \text{for } x > 0, \quad (19)$$

where  $\phi(x) := \Phi'(x) = \exp(-x^2/2)/\sqrt{2\pi}$ . In particular, as  $x \rightarrow \infty$ ,

$$\begin{aligned} \Phi(-x) &= \exp(-x^2/2 + O(\log x)) \quad \text{and} \\ C(\Phi(x)) &= \log(O(1) + \log(1/\Phi(-x))) = 2\log(x) - \log(2) + o(1). \end{aligned}$$

Now consider two sequences  $(x_n)_n$  and  $(\mu_n)_n$  tending to  $\infty$  and  $F_o = \Phi$ ,  $F_n = (1 - \varepsilon_n)\Phi + \varepsilon_n\Phi(\cdot - \mu_n)$ . Then the inequalities (19) imply that

$$\begin{aligned} \Gamma(F_o(x_n))F_o(x_n)(1 - F_o(x_n)) &= (2\log(x_n) + O(1))\Phi(-x_n)(1 + o(1)) \\ &= \exp(-x_n^2/2 + O(\log(x_n))). \end{aligned}$$

Moreover,

$$F_o(x_n) - F_n(x_n) = \varepsilon_n(\Phi(\mu_n - x_n) - \Phi(-x_n)) = \varepsilon_n\Phi(\mu_n - x_n)(1 + o(1)),$$

because  $\Phi(-x_n) \leq \phi(x_n)/x_n$  while

$$\Phi(\mu_n - x_n) \geq \begin{cases} 1/2 & \text{if } \mu_n \geq x_n, \\ \frac{\phi(x_n - \mu_n)}{x_n - \mu_n + 1} \geq \frac{\phi(x_n) \exp(\mu_n^2/2)}{x_n + 1} & \text{if } \mu_n < x_n. \end{cases}$$

Consequently,  $\Delta_n(F_n, \Phi) \rightarrow \infty$  if

$$\frac{n\varepsilon_n\Phi(\mu_n - x_n)}{n^{1/2} \exp(-x_n^2/4 + O(\log(x_n))) + O(\log(x_n))} \rightarrow \infty. \quad (20)$$

In part (a) with  $\varepsilon_n = n^{-\beta+o(1)}$  and  $\beta \in (1/2, 1)$  we imitate the arguments of Donoho and Jin (2004) and consider

$$\mu_n = \sqrt{2r \log(n)} \quad \text{and} \quad x_n = \sqrt{2q \log(n)}$$

with  $0 < r < q \leq 1$ . Then by (19),

$$\begin{aligned} n\varepsilon_n\Phi(\mu_n - x_n) &= n^{1-\beta-(\sqrt{q}-\sqrt{r})^2+o(1)}, \\ n^{1/2} \exp(-x_n^2/4 + O(\log(x_n))) &= n^{1/2-q/2+o(1)}, \\ O(\log(x_n)) &= n^{o(1)}, \end{aligned}$$

so the left hand side of (20) equals

$$\frac{n^{1-\beta-(\sqrt{q}-\sqrt{r})^2+o(1)}}{n^{1/2-q/2+o(1)} + n^{o(1)}} = \frac{n^{1/2-\beta+q/2-(\sqrt{q}-\sqrt{r})^2+o(1)}}{1 + n^{(q-1)/2+o(1)}} = \frac{n^{1/2-\beta+2\sqrt{r}\sqrt{q}-\sqrt{q}^2/2-r+o(1)}}{1 + n^{(q-1)/2+o(1)}}.$$

The exponent in the enumerator is maximal in  $q \in (r, 1]$  if  $\sqrt{q} = \min\{2\sqrt{r}, 1\}$ , i.e.  $q = \min\{4r, 1\}$ , and this leads to

$$\begin{cases} 1/2 - \beta + r & \text{if } r \leq 1/4, \\ 1 - \beta - (1 - \sqrt{r})^2 & \text{if } r \geq 1/4. \end{cases}$$

Thus when  $\beta \in (1/2, 3/4)$  we should choose  $r \in (\beta - 1/2, 1/4)$  and  $q = 4r$ . When  $\beta \in [3/4, 1)$  we should choose  $r \in \left((1 - \sqrt{1 - \beta})^2, 1\right)$  and  $q = 1$ .

As to part (b), we consider the more general setting that  $\varepsilon_n = n^{-\beta+o(1)}$  for some  $\beta \in [1/2, 3/4)$ , where  $\pi_n = \sqrt{n}\varepsilon_n \rightarrow 0$ . The latter constraint is trivial when  $\beta > 1/2$  but relevant when  $\beta = 1/2$ . Now we consider

$$\mu_n := \sqrt{2s \log(1/\pi_n)} \quad \text{and} \quad x_n := \sqrt{2q \log(1/\pi_n)}$$

with arbitrary constants  $0 < s < q$ . Now

$$\begin{aligned} n\varepsilon_n \Phi(\mu_n - x_n) &= n^{1/2} \pi_n \Phi(\mu_n - x_n) \\ &= n^{1/2} \pi_n^{1+(\sqrt{q}-\sqrt{s})^2+o(1)}, \\ n^{1/2} \exp(-x_n^2/4 + O(\log(x_n))) &= n^{1/2} \pi_n^{q/2+o(1)}, \\ O(\log(x_n)) &= \pi_n^{o(1)}, \end{aligned}$$

so the left hand side of (20) equals

$$\frac{n^{1/2} \pi_n^{1+(\sqrt{q}-\sqrt{s})^2+o(1)}}{n^{1/2} \pi_n^{q/2+o(1)} + \pi_n^{o(1)}} = \frac{\pi_n^{1+\sqrt{q}^2/2-2\sqrt{q}\sqrt{s}+s+o(1)}}{1 + n^{-1/2} \pi_n^{-q/2+o(1)}} = \frac{\pi_n^{1+\sqrt{q}^2/2-2\sqrt{q}\sqrt{s}+s+o(1)}}{1 + n^{-1/2+( \beta-1/2)q/2+o(1)}}.$$

The exponent of  $\pi_n$  becomes minimal in  $q \in (s, \infty)$  if  $\sqrt{q} = 2\sqrt{s}$ , i.e.  $q = 4s$ . Then we obtain

$$\frac{\pi_n^{1-s+o(1)}}{1 + n^{-1/2+(2\beta-1)s+o(1)}} = \frac{\pi_n^{1-s+o(1)}}{1 + \sqrt{n}^{(4\beta-2)s-1+o(1)}},$$

and this converges to  $\infty$  if the exponents of  $\pi_n$  and  $\sqrt{n}$  are negative and non-positive, respectively.

This is the case if  $1 < s \leq 1/(4\beta - 2)$ . (Note that  $4\beta - 2 < 1$  because  $\beta < 3/4$ .)  $\square$

**Proof of Theorem 5.1.** By symmetry it suffices to analyze the differences  $b_{nj} - s_{nj}$  and  $b_{nj}^{\text{BJO}} - s_{nj}$  for  $0 \leq j < n$ .

Recall the notation  $b(s, \gamma)$  for the unique number  $b \in (s, 1)$  such that  $K(s, b) = \gamma$ , introduced in (K.6). There we considered only  $s \in (0, 1)$ , but it follows from  $K(0, b) = -\log(1 - b)$  that  $b(0, \gamma) = 1 - \exp(-\gamma) = \gamma + o(1)$  as  $\gamma \rightarrow 0$ . For  $0 \leq j < n$ , we may write

$$b_{nj}^{\text{BJO}} = b(s_{nj}, \gamma_n^{\text{BJO}}) \quad \text{and} \quad b_{nj} \leq b(t_{n,j+1}, \gamma_n(t_{n1}))$$

Recall that

$$\gamma_n^{\text{BJ}} = \frac{\log \log n}{n}(1 + o(1)) \quad \text{and} \quad \gamma_n(t_{n1}) = \frac{\log \log n}{n}(1 + o(1)).$$

Moreover, since  $K(s_{nj}, \cdot)$  is convex on  $[s_{nj}, 1)$ , the numbers  $b(s_{nj}, \gamma)$  are concave in  $\gamma \geq 0$ . In particular, with  $\tilde{\gamma}_n$  denoting the maximum of  $\gamma_n^{\text{BJO}}$  and  $\gamma_n(t_{n1})$ ,

$$\frac{b_{nj}^{\text{BJO}} - s_{nj}}{b(s_{nj}, \tilde{\gamma}_n) - s_{nj}} \geq \frac{\gamma_n^{\text{BJ}}}{\tilde{\gamma}_n} \rightarrow 1$$

uniformly in  $0 \leq j < n$ . Hence it suffices to show that

$$\limsup_{n \rightarrow \infty} \max_{0 \leq j < n} \frac{b(t_{n,j+1}, \tilde{\gamma}_n) - s_{nj}}{b(s_{nj}, \tilde{\gamma}_n) - s_{nj}} \leq 1. \quad (21)$$

First we consider indices  $j \leq j(n, 1) := \lceil (\log \log n)^{1/2} \rceil$ . Note that for  $j = 0$ ,

$$b(s_{nj}, \tilde{\gamma}_n) - s_{nj} = b(0, \tilde{\gamma}_n) = \tilde{\gamma}_n(1 + o(1)),$$

and we may deduce from (13) and  $\lim_{y \rightarrow \infty} H^{-1}(y)/y = 1$  that uniformly in  $1 \leq j \leq j(n, 1)$ ,

$$\begin{aligned} b(s_{nj}, \tilde{\gamma}_n) - s_{nj} &\geq (1 + o(1))s_{nj}H^{-1}(\tilde{\gamma}_n/s_{nj}) \\ &= (1 + o(1))\tilde{\gamma}_n \frac{H^{-1}(n\tilde{\gamma}_n/j)}{n\tilde{\gamma}_n/j} \\ &\geq \tilde{\gamma}_n(1 + o(1)). \end{aligned}$$

On the other hand, since

$$t_{n,j+1} - s_{nj} = \frac{1 - s_{nj}}{n + 1} < n^{-1} = o(\tilde{\gamma}_n),$$

we may conclude that uniformly in  $0 \leq j \leq j(n, 1)$ ,

$$\begin{aligned} b(t_{n,j+1}, \tilde{\gamma}_n) - s_{nj} &\leq b(t_{n,j+1}, \tilde{\gamma}_n) - t_{n,j+1} + n^{-1} \\ &\leq t_{n,j+1}H^{-1}(\tilde{\gamma}_n/t_{n,j+1}) + n^{-1} \\ &= \tilde{\gamma}_n \frac{H^{-1}(\tilde{\gamma}_n/t_{n,j+1})}{\tilde{\gamma}_n/t_{n,j+1}} + n^{-1} \\ &\leq \tilde{\gamma}_n(1 + o(1)). \end{aligned}$$

Hence (21) holds true if we restrict  $j$  to the interval  $\{0, \dots, j(n, 1)\}$ .

Next we consider indices  $j$  between  $j(n, 1)$  and  $j(n, 2) := \lceil n\tilde{\gamma}_n^{1/3} \rceil$ , i.e.  $j(n, 2)/n \rightarrow 0$  and  $t_{n,j+1}/s_{nj} \rightarrow 1$  uniformly in  $j(n, 1) \leq j \leq j(n, 2)$ . Then it follows from (13), together with

$H^{-1}(y) \geq y$  and monotonicity of  $H^{-1}(\cdot)$ , that uniformly in  $j_{n1} \leq j \leq j_{n2}$ ,

$$\begin{aligned} \frac{b(t_{n,j+1}, \tilde{\gamma}_n) - s_{nj}}{b(s_{nj}, \tilde{\gamma}_n) - s_{nj}} &= (1 + o(1)) \frac{t_{n,j+1} H^{-1}(\tilde{\gamma}_n/t_{n,j+1}) + n^{-1}}{s_{nj} H^{-1}(\tilde{\gamma}_n/s_{nj})} \\ &= (1 + o(1)) \frac{t_{n,j+1} H^{-1}(\tilde{\gamma}_n/t_{n,j+1})}{s_{nj} H^{-1}(\tilde{\gamma}_n/s_{nj})} \\ &\leq (1 + o(1)) \frac{t_{n,j+1}}{s_{nj}} \\ &= 1 + o(1). \end{aligned}$$

Hence (21) is satisfied with  $\{j(n, 1), \dots, j(n, 2)\}$  in place of  $\{0, 1, \dots, n-1\}$ .

Now consider  $j(n, 3) := n - j(n, 2)$ . Uniformly in  $j(n, 2) \leq j \leq j(n, 3)$ , the product  $s_{nj}(1 - s_{nj})$  is larger than  $\tilde{\gamma}_n^{1/3}(1 + o(1))$ , so  $\tilde{\gamma}_n/(s_{nj}(1 - s_{nj})) \rightarrow 0$ . Moreover,  $\text{logit}(t_{n,j+1}) - \text{logit}(s_{nj}) \rightarrow 0$ , and it follows from (14) that

$$\begin{aligned} \frac{b(t_{n,j+1}, \tilde{\gamma}_n) - s_{nj}}{b(j/n, \tilde{\gamma}_n) - s_{nj}} &\leq (1 + o(1)) \frac{\sqrt{2\tilde{\gamma}_n t_{n,j+1}(1 - t_{n,j+1})} + n^{-1}}{\sqrt{2\tilde{\gamma}_n s_{nj}(1 - s_{nj})}} \\ &= 1 + o(1) + O(\tilde{\gamma}_n^{-1/3} n^{-1}) \\ &= 1 + o(1) \end{aligned}$$

uniformly in  $j(n, 2) \leq j \leq j(n, 3)$ .

Finally, we may conclude from (15), concavity of  $\tilde{H}^{-1}(\cdot)$  and the inequality  $H^{-1}(y) \geq 1 - e^{-y}$  that that uniformly for  $j(n, 3) \leq j \leq n-1$ ,

$$\begin{aligned} \frac{b(t_{n,j+1}, \tilde{\gamma}_n) - s_{nj}}{b(j/n, \tilde{\gamma}_n) - s_{nj}} &\leq (1 + o(1)) \frac{(1 - t_{n,j+1}) \tilde{H}^{-1}(\tilde{\gamma}_n/(1 - t_{n,j+1})) + (1 - s_{nj})/n}{(1 - s_{nj}) \tilde{H}^{-1}(\tilde{\gamma}_n/(1 - s_{nj}))} \\ &\leq (1 + o(1)) \left( 1 + \frac{(1 - s_{nj})/n}{(1 - t_{n,j+1}) \tilde{H}^{-1}(\tilde{\gamma}_n/(1 - t_{n,j+1}))} \right) \\ &= (1 + o(1)) (1 + O(n^{-1} \tilde{\gamma}_n^{-2/3})) \\ &= 1 + o(1). \end{aligned}$$

These considerations prove (21).

It remains to analyze the maximum of  $b_{nj}^{\text{BJO}} - s_{nj}$  and  $b_{nj} - s_{nj}$ , respectively, over  $j = 0, 1, \dots, n$ . Note first that by (K.5),

$$b_{nj}^{\text{BJO}} - s_{nj} \leq \sqrt{2\tilde{\gamma}_n s_{nj}(1 - s_{nj})} + \tilde{\gamma}_n \leq \sqrt{\tilde{\gamma}_n/2} + \tilde{\gamma}_n = (1 + o(1)) \sqrt{\frac{\log \log n}{2n}}.$$

On the other hand, for  $j(n) := \lfloor (n+1)/2 \rfloor$ , (14) implies that

$$b_{n,j(n)}^{\text{BJO}} - s_{n,j(n)} \geq (1 + o(1)) \sqrt{2\gamma_n^{\text{BJ}} s_{n,j(n)}(1 - s_{n,j(n)})} = (1 + o(1)) \sqrt{\frac{\log \log n}{2n}}.$$

This proves the assertion about  $\max_j (b_{nj}^{\text{BJO}} - s_{nj})$ . As to the new confidence bounds, note first that by (K.5),

$$\begin{aligned} b_{nj} - s_{nj} &\leq b_{nj} - t_{nj} + n^{-1} \\ &\leq \sqrt{2\gamma_n(t_{nj})t_{nj}(1-t_{nj})} + \gamma_n(t_{n1}) + n^{-1} \\ &\leq n^{-1/2} \sqrt{h(t_{nj}) + 2t_{nj}(1-t_{nj})\tilde{\kappa}_{n,\nu,\alpha}} + O(n^{-1} \log \log n), \end{aligned}$$

where  $h(t) := 2t(1-t)(C(t) + \nu D(t))$  is a continuous function on  $(0, 1)$  with limit 0 as  $t \rightarrow \{0, 1\}$ . Consequently,  $\sup_{(0,1)} h$  is finite and

$$\max_{j=0,1,\dots,n} (b_{nj} - s_{nj}) \leq n^{-1/2} \sqrt{\sup_{(0,1)} h + \tilde{\kappa}_{n,\nu,\alpha}/2} + O(n^{-1} \log \log n) = O(n^{-1/2}).$$

□

**Acknowledgements.** We are grateful to Guenther Walther for stimulating talks about likelihood ratio tests in nonparametric settings. In particular, Rivera and Walther (2013) inspired us to reformulate the Law of the Iterated Logarithm in terms of processes with subexponential tails. Many thanks go to Rudy Beran for drawing our attention to Bahadur and Savage (1956).

## References

- [1] S. ALDOR-NOIMAN, L. D. BROWN, A. BUJA, W. ROLKE, AND R. A. STINE, *The power to see: A new graphical test of normality*, Amer. Statist., 67 (2013), pp. 249–260.
- [2] R. R. BAHADUR AND L. J. SAVAGE, *The nonexistence of certain statistical procedures in nonparametric problems*, Ann. Math. Statist., 27 (1956), pp. 1115–1122.
- [3] R. H. BERK AND D. H. JONES, *Goodness-of-fit test statistics that dominate the Kolmogorov statistics*, Z. Wahrsch. Verw. Gebiete, 47 (1979), pp. 47–59.
- [4] D. DONOHO AND J. JIN, *Higher criticism for detecting sparse heterogeneous mixtures*, Ann. Statist., 32 (2004), pp. 962–994.
- [5] L. DÜMBGEN, *New goodness-of-fit tests and their application to nonparametric confidence sets*, Ann. Statist., 26 (1998), pp. 288–314.
- [6] A. M. EIGER, B. NADLER, AND C. SPIEGELMAN, *The calibrated kolmogorov-smirnov test*, tech. rep., Department of Computer Science, Weizmann Institute of Science; Department of Statistics, Texas A&M University, 2013. (<http://arxiv.org/abs/1311.3190>)

- [7] P. ERDÖS, *On the law of the iterated logarithm*, Ann. of Math. (2), 43 (1942), pp. 419–436.
- [8] W. HOEFFDING, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc., 58 (1963), pp. 13–30.
- [9] K. ITÔ AND H. P. MCKEAN, JR., *Diffusion processes and their sample paths*, Springer-Verlag, Berlin, 1974. Second printing, corrected, Die Grundlehren der mathematischen Wissenschaften, Band 125.
- [10] L. JAGER AND J. A. WELLNER, *Goodness-of-fit tests via phi-divergences*, Ann. Statist., 35 (2007), pp. 2018–2053.
- [11] J. KIEFER, *Iterated logarithm analogues for sample quantiles when  $p_n \downarrow 0$* , in Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, University of California, 1973, pp. 227–244.
- [12] L. LE CAM AND G. L. YANG, *Asymptotics in statistics – Some basic concepts*, Springer Series in Statistics, Springer-Verlag, New York, second ed., 2000.
- [13] D. M. MASON AND J. H. SCHUENEMEYER, *A modified Kolmogorov-Smirnov test sensitive to tail alternatives*, Ann. Statist., 11 (1983), pp. 933–946.
- [14] A. B. OWEN, *Nonparametric likelihood confidence bands for a distribution function*, J. Amer. Statist. Assoc., 90 (1995), pp. 516–521.
- [15] C. RIVERA AND G. WALTHER, *Optimal detection of a jump in the intensity of a Poisson process or in a density with likelihood ratio statistics*, Scand. J. Statist., 40 (2013), pp. 752–769.
- [16] G. R. SHORACK AND J. A. WELLNER, *Empirical processes with applications to statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1986.

## 7 Supplementary material

### 7.1 A remark on moment-generating functions

Somewhat hidden in our proofs of Lemmas 3.1 and 3.3 is a basic fact about moment generating functions which is stated in a slightly different form by Rivera and Walther (2013) and possibly of independent interest: Suppose that  $X$  is a real-valued random variable with mean  $\mu$  and moment-generating function  $m_X$ ,

$$m_X(t) := \mathbb{E} \exp(tX).$$

We assume that  $m_X < \infty$  in a neighborhood of zero. In particular, all moments of  $X$  are finite. A standard application of Markov's inequality yields

$$\mathbb{P}(X \geq x) \leq \exp(-K(x)) \quad \text{for all } x \geq \mu,$$

$$\mathbb{P}(X \leq x) \leq \exp(-K(x)) \quad \text{for all } x \leq \mu,$$

where

$$K(x) := \sup_{t \in \mathbb{R}} (tx - \log m_X(t)) \begin{cases} = \sup_{t \geq 0} (tx - \log m_X(t)) & \text{if } x \geq \mu, \\ = \sup_{t \leq 0} (tx - \log m_X(t)) & \text{if } x \leq \mu. \end{cases}$$

The latter facts follow from the fact that  $\log m_X$  is a convex function with derivative  $\mu$  at 0. Note also that  $K : \mathbb{R} \rightarrow [0, \infty]$  is a convex, lower semi-continuous function with  $K(\mu) = 0$  and  $\lim_{|x| \rightarrow \infty} K(x) = \infty$ . From this one can derive the following inequalities:

**Lemma 7.1.** *For arbitrary  $\eta > 0$ ,*

$$\left. \begin{aligned} &\mathbb{P}(K(X) \geq \eta \text{ and } X \geq \mu) \\ &\mathbb{P}(K(X) \geq \eta \text{ and } X \leq \mu) \end{aligned} \right\} \leq \exp(-\eta),$$

and thus

$$\mathbb{P}(K(X) \geq \eta) \leq 2 \exp(-\eta).$$

**Proof of Lemma 7.1.** By symmetry, it suffices to show that  $\mathbb{P}(K(X) \geq \eta \text{ and } X \geq \mu)$  is not greater than  $\exp(-\eta)$ . Since  $K : [\mu, \infty) \rightarrow [0, \infty]$  is convex and lower semi-continuous with  $K(\mu) = 0$  and  $\lim_{x \rightarrow \infty} K(x) = \infty$ , the point

$$x_\eta := \max\{x \geq \mu : K(x) \leq \eta\}$$

is well-defined. When  $K(x_\eta) = \eta$ , convexity of  $K$  and  $K(\mu) = 0$  imply that  $K(x) < \eta$  for all  $x \in [\mu, x_\eta)$ . Hence

$$\mathbb{P}(K(X) \geq \eta \text{ and } X \geq \mu) = \mathbb{P}(X \geq x_\eta) \leq \exp(-K(x_\eta)) = \exp(-\eta).$$



When  $K(x_\eta) < \eta$ , we may conclude from monotonicity and lower semicontinuity of  $K$  that  $K(x) = \infty$  for all  $x > x_\eta$ . But this implies that

$$\mathbb{P}(K(X) \geq \eta \text{ and } X \geq \mu) = \mathbb{P}(X > x_\eta) = \sup_{x > x_\eta} \mathbb{P}(X \geq x) = 0.$$

□

## 7.2 Exponential inequalities for beta distributions

Let  $s, t \in (0, 1)$ , and let  $Y \sim \text{Beta}(mt, m(1-t))$  for some  $m > 0$ . Then

$$\begin{aligned} \mathbb{P}(Y \leq s) &\leq \inf_{\lambda \leq 0} \mathbb{E} \exp(\lambda Y - \lambda s) \leq \exp(-mK(t, s)) \quad \text{if } s \leq t, \\ \mathbb{P}(Y \geq s) &\leq \inf_{\lambda \geq 0} \mathbb{E} \exp(\lambda Y - \lambda s) \leq \exp(-mK(t, s)) \quad \text{if } s \geq t. \end{aligned}$$

**Proof.** In case of  $s \geq t$ , Markov's inequality yields that

$$\mathbb{P}(Y \geq s) = \inf_{\lambda \geq 0} \mathbb{P}(\lambda Y - \lambda s \geq 0) \leq \inf_{\lambda \geq 0} \mathbb{E} \exp(\lambda Y - \lambda s) = \inf_{\lambda \geq 0} \mathbb{E} \exp(\lambda m Y - \lambda m s).$$

The latter step is trivial but convenient for the next consideration: We may write  $Y = G/(G + G')$  with independent random variables  $G \sim \text{Gamma}(mt)$  and  $G' \sim \text{Gamma}(m(1-t))$ . Moreover, it is well-known that  $Y$  and  $G + G'$  are stochastically independent with  $\mathbb{E}(G + G') = m$ . Consequently, by Jensen's inequality and Fubini's theorem,

$$\begin{aligned} \mathbb{E} \exp(\lambda m Y - \lambda m s) &= \mathbb{E} \exp(\lambda \mathbb{E}(G - s(G + G') \mid Y)) \\ &= \mathbb{E} \exp(\lambda \mathbb{E}((1-s)G - \lambda s G' \mid Y)) \\ &\leq \mathbb{E} \mathbb{E}(\exp(\lambda(1-s)G - \lambda s G') \mid Y) \\ &= \mathbb{E} \exp(\lambda(1-s)G - \lambda s G') \\ &= \mathbb{E} \exp(\lambda(1-s)G) \mathbb{E} \exp(-\lambda s G') \\ &= (1 - \lambda(1-s))^{-mt} (1 + \lambda s)^{-m(1-t)} \\ &= \exp\left(-m(t \log(1 - \lambda(1-s)) + (1-t) \log(1 + \lambda s))\right) \end{aligned}$$

for  $0 \leq \lambda < 1/(1-s)$ . (For  $\lambda \geq 1/(1-s)$  the expectation of  $\exp(\lambda(1-s)G)$  would be infinite.) Elementary calculations show that  $t \log(1 - \lambda(1-s)) + (1-t) \log(1 + \lambda s)$  is maximal for  $\lambda = (s-t)/(s(1-s)) \in [0, 1/(1-s))$ , and this yields the bound

$$\inf_{\lambda \geq 0} \mathbb{E} \exp(\lambda Y - \lambda s) \leq \exp(-mK(t, s)).$$

In case of  $s \leq t$ , the previous result may be applied to  $1 - Y \sim \text{Beta}(m(1 - t), mt)$ :

$$\begin{aligned} \mathbb{P}(Y \leq s) &= \mathbb{P}(1 - Y \geq 1 - s) \leq \inf_{\lambda \geq 0} \mathbb{E} \exp(\lambda(1 - Y) - \lambda(1 - s)) \\ &\begin{cases} = \inf_{\lambda \leq 0} \mathbb{E} \exp(\lambda Y - \lambda s), \\ \leq \exp(-mK(1 - t, 1 - s)) = \exp(-mK(t, s)). \end{cases} \quad \square \end{aligned}$$

### 7.3 Further details about Gaussian mixtures

As in Section 4 we consider the standard Gaussian distribution function  $\Phi$  and the alternative distribution functions

$$F_n := (1 - \varepsilon_n)\Phi + \varepsilon_n\Phi(\cdot - \mu_n),$$

where  $\varepsilon_n \downarrow 0$  and  $\mu_n \rightarrow \infty$ . Optimal tests of  $H_0 : F \equiv \Phi$  versus  $H_1 : F \equiv F_n$  reject for large values of the log-likelihood ratio statistic

$$\sum_{i=1}^n \log \frac{dF_n}{d\Phi}(X_i) = \sum_{i=1}^n \log(1 + V_n(X_i))$$

with

$$V_n(x) = \varepsilon_n(\exp(\mu_n x - \mu_n^2/2) - 1).$$

If  $(\mu_n)_n$  is chosen such that

$$\sum_{i=1}^n \log(1 + V_n(X_i)) \rightarrow_p 0 \quad \text{when } F \equiv \Phi, \quad (22)$$

then for any sequence of tests  $\phi_n : \mathbb{R}^n \rightarrow [0, 1]$ ,

$$\limsup_{n \rightarrow \infty} |\mathbb{E}_{F_n} \phi_n(X_1, \dots, X_n) - \mathbb{E}_{\Phi} \phi_n(X_1, \dots, X_n)| = 0;$$

see LeCam and Yang (2000).

**Lemma 7.2.** *Suppose that  $\varepsilon_n = n^{-\beta+o(1)}$  for some  $\beta \in [1/2, 3/4]$  and  $\pi_n = n^{1/2}\varepsilon_n \rightarrow 0$ . Then (22) is satisfied if  $\mu_n = \sqrt{2s \log(1/\pi_n)}$  for some fixed  $s \in (0, 1)$ .*

**Proof of Lemma 7.2.** Note that for  $v > -1$ ,

$$\log(1 + v) = v - \frac{v^2}{2(1 + \xi(v))}$$

with  $\xi(v) \geq \min\{0, v\}$ . Consequently, since  $V_n > -\varepsilon_n$ ,

$$\sum_{i=1}^n V_n(X_i) - \frac{1}{2(1 - \varepsilon_n)} \sum_{i=1}^n V_n(X_i)^2 \leq \sum_{i=1}^n \log(1 + V_n(X_i)) \leq \sum_{i=1}^n V_n(X_i).$$

But it follows from  $\mathbb{E}_\Phi(V_n(X_1)) = 0$  that

$$\begin{aligned}
\mathbb{E}_\Phi\left(\left(\sum_{i=1}^n V_n(X_i)\right)^2\right) &= n \operatorname{Var}_\Phi(V_n(X_1)) \\
&= n\varepsilon_n^2 (\mathbb{E}_\Phi \exp(2\mu_n X_1 - \mu_n^2) - 1) \\
&= \pi_n^2 (\exp(\mu_n^2) - 1) \\
&= \pi_n^{2(1-s)} - \pi_n^2 \\
&\rightarrow 0
\end{aligned}$$

because  $s < 1$ , and

$$\mathbb{E}_\Phi\left(\frac{1}{2(1-\varepsilon_n)} \sum_{i=1}^n V_n(X_i)^2\right) = \frac{n \operatorname{Var}_\Phi(V_n(X_1))}{2(1-\varepsilon_n)} \rightarrow 0.$$

□

#### 7.4 Bahadur and Savage (1956) revisited

Let  $(L_n, U_n)$  be a  $(1 - \alpha)$ -confidence band for  $F \in \mathcal{F}$  with a given class  $\mathcal{F}$  of distribution functions. That means  $L_n = L_n(\cdot, \mathbf{X}_n)$  and  $U_n = U_n(\cdot, \mathbf{X}_n)$  are non-decreasing functions on the real line depending on the data vector  $\mathbf{X}_n = (X_i)_{i=1}^n$  such that

$$\mathbb{P}_F(L_n \leq F \leq U_n \text{ on } \mathbb{R}) \geq 1 - \alpha \quad \text{for any } F \in \mathcal{F}.$$

We assume that  $\mathcal{F}$  is convex and satisfies  $F(\cdot - \mu) \in \mathcal{F}$  for any  $F \in \mathcal{F}$  and  $\mu \in \mathbb{R}$ . This is true if, for instance,  $\mathcal{F}$  corresponds to all mixtures of Gaussian distributions with variance one. Then Theorem 2 of Bahadur and Savage (1956) may be modified as follows:

**Theorem 7.3.** *Let  $(L_n, U_n)$  be a  $(1 - \alpha)$ -confidence band for  $F \in \mathcal{F}$ . For any  $\varepsilon \in (0, 1)$ ,*

$$\begin{aligned}
\mathbb{P}_F\left(\inf_{x \in \mathbb{R}} U_n(x) < \varepsilon\right) &\leq (1 - \varepsilon)^{-n} \alpha, \\
\mathbb{P}_F\left(\sup_{x \in \mathbb{R}} L_n(x) > 1 - \varepsilon\right) &\leq (1 - \varepsilon)^{-n} \alpha.
\end{aligned}$$

Setting  $\varepsilon = c/n$  for some fixed  $c > 0$  reveals that  $\inf_{x \in \mathbb{R}} U_n(x) < c/n$  or  $\sup_{x \in \mathbb{R}} L_n(x) \leq 1 - c/n$  with probability at most  $(1 - c/n)^{-n} \alpha = e^c \alpha + o(1)$ , respectively.

**Proof of Theorem 7.3.** By symmetry, it suffices to prove the claim about  $U_n$ . By monotonicity of  $U_n$ ,

$$\mathbb{P}_F\left(\inf_{x \in \mathbb{R}} U_n(x) < \varepsilon\right) = \sup_{x \in \mathbb{R}, \delta \in (0, \varepsilon)} \mathbb{P}_F(U_n(x) < \delta).$$

Hence it suffices to show that  $\mathbb{P}_F(U_n(x) < \delta) \leq (1 - \varepsilon)^{-n} \alpha$  for any single point  $x \in \mathbb{R}$  and  $\delta \in (0, \varepsilon)$ . To this end consider  $F_{\varepsilon, \mu} := (1 - \varepsilon)F + \varepsilon F(\cdot - \mu)$  for our given  $\varepsilon$  and some  $\mu \in \mathbb{R}$ . Note that  $\mathcal{L}_{F_{\varepsilon, \mu}}(\mathbf{X}_n)$  describes the distribution of

$$\tilde{\mathbf{X}}_n := (Y_i + \xi_i \mu)_{i=1}^n$$

with  $2n$  independent random variables  $\xi_1, \xi_2, \dots, \xi_n \sim \text{Bin}(1, \varepsilon)$  and  $Y_1, Y_2, \dots, Y_n \sim F$ . In particular, for any event  $A_n \subset \mathbb{R}^n$ ,

$$\begin{aligned} \mathbb{P}_{F_{\varepsilon, \mu}}(\mathbf{X}_n \in A_n) &= \mathbb{P}(\tilde{\mathbf{X}}_n \in A_n) \\ &\geq \mathbb{P}(\tilde{\mathbf{X}}_n \in A_n, \xi_1 = \xi_2 = \dots = \xi_n = 0) \\ &= (1 - \varepsilon)^n \mathbb{P}_F(\mathbf{X}_n \in A_n). \end{aligned}$$

Consequently, since  $F_{\varepsilon, \mu} \in \mathcal{F}$ , too, we may conclude from

$$\mathbb{P}_{F_{\varepsilon, \mu}}(L_n \leq F_{\varepsilon, \mu} \leq U_n \text{ on } \mathbb{R}) \geq 1 - \alpha$$

that

$$\begin{aligned} \alpha &\geq \mathbb{P}_{F_{\varepsilon, \mu}}(U_n(x) < F_{\varepsilon, \mu}(x)) \\ &\geq (1 - \varepsilon)^n \mathbb{P}_F(U_n(x) < (1 - \varepsilon)F(x) + \varepsilon F(x - \mu)) \\ &\geq (1 - \varepsilon)^n \mathbb{P}_F(U_n(x) < \varepsilon F(x - \mu)). \end{aligned}$$

But for sufficiently small (negative)  $\mu$ , the value  $\varepsilon F(x - \mu)$  is greater than or equal to  $\delta$ . Then we may conclude that  $\alpha \geq (1 - \varepsilon)^n \mathbb{P}_F(U_n(x) < \delta)$ .  $\square$